

UNIVERSIDAD CARLOS III DE MADRID

ESCUELA POLITÉCNICA SUPERIOR

**INGENIERÍA TÉCNICA DE TELECOMUNICACIÓN
SISTEMAS DE TELECOMUNICACIÓN**



PROYECTO FINAL DE CARRERA

**ALGORITMOS UNIVARIANTES Y
MULTIVARIANTES PARA LA GENERACIÓN
DE MAPAS DE ACTIVACIÓN EN RESONANCIA
MAGNÉTICA FUNCIONAL**

**AUTOR: LETICIA TALAVERA CARRASCO
TUTOR: MANEL MARTÍNEZ RAMÓN**

ENERO 2011

TÍTULO: “*ALGORITMOS UNIVARIANTES Y MULTIVARIANTES PARA LA GENERACIÓN DE MAPAS DE ACTIVACIÓN EN RESONANCIA MAGNÉTICA FUNCIONAL*”.

AUTOR: *Leticia Talavera Carrasco*

TUTOR: *Manel Martínez Ramón*

La defensa del presente Proyecto Fin de Carrera se realizó el día 20 de Enero de 2011; siendo calificada por el siguiente tribunal:

PRESIDENTE: *Dr. Harold Molina Bulla*

SECRETARIO *Dr. José Miguel Leiva Murillo*

VOCAL *Dr. Ricardo Vergaz Benito*

Habiendo obtenido la siguiente calificación:

CALIFICACIÓN:

Presidente

Secretario

Vocal

Agradecimientos

A toda mi familia, en especial a mis padres (Francisco Javier y Alfonsa), que gracias a su apoyo, su confianza y su cariño han hecho posible que haya llegado a la culminación de esta etapa universitaria.

A Carlos, una persona muy especial, por comprenderme en lo incomprensible. Por apoyarme y animarme a seguir adelante, porque con él todo ha sido más fácil.

A mis compañeros de la carrera, por su ayuda y apoyo prestados. De los cuales me llevo buenos amigos: Mónica, Miriam, Fernando, Rosana...y muy buenos momentos.

A mis amigos de Fuenlabrada por hacerme reír y ayudarme en los malos momentos.

A Manel, tutor de mi proyecto, por su paciencia y ayuda prestada.

En definitiva, a toda la gente que ha compartido conmigo estos años universitarios.

Gracias a todos. Nunca os olvidaré.

Leticia Talavera Carrasco

*Si lloras por haber perdido el sol,
las lágrimas no te dejarán ver las estrellas.*

Rabindranath Tagore

*Nuestra recompensa se encuentra en el esfuerzo y no en el resultado,
un esfuerzo total es una victoria completa.*

Mahatma Gandhi

*Si una persona es perseverante, aunque sea dura de entendimiento,
se hará inteligente; y aunque sea débil se transformará en fuerte.*

Leonardo Da Vinci

*A veces navegamos con el viento, a veces en contra;
pero debemos navegar, no estar a la deriva, ni echar el ancla.*

Oliver Wendell Holmes

Resumen

Este proyecto tiene una aplicación en el campo de la ingeniería biomédica.

A partir de datos fMRI (imágenes de Resonancia Magnética Funcional) se programan y prueban una serie de algoritmos univariantes y multivariantes para la generación de mapas de activación y así poder indicar de forma precisa las zonas del cerebro que son importantes en la realización de alguna tarea específica. En este proyecto se usan datos de activación motora.

Los algoritmos univariantes utilizan el Modelo Lineal General (GLM) para la generación de mapas de activación cerebral.

Los algoritmos multivariantes utilizan dos variantes para el cálculo de mapas de activación. Una de ellas está basada en Máquinas de Vectores Soporte (SVM) y la otra en Máquinas de Procesos Gaussianos (GP).

Además, se prueban dos tipos de matrices de covarianzas, la lineal y la no lineal.

ÍNDICE GENERAL

Índice de figuras	13
Índice de tablas	19
1. Resonancia Magnética Funcional	21
1.1. Introducción	21
1.2. Principios físicos de la resonancia magnética	21
1.3. Bases neurofisiológicas de la resonancia magnética funcional	34
1.4. Diseño de tareas para Resonancia Magnética Funcional	40
1.5. Preprocesado en Resonancia Magnética Funcional	50
1.6. Métodos univariantes: GLM	56
1.6.1. Modelo de datos	56
1.6.2. Mapa paramétrico estadístico (SPM)	57
2. Algoritmos multivariantes: SVM y Redes de Procesos Gaussianos	59
2.1. Máquinas de Vectores Soporte (SVM)	59
2.1.1. Introducción	59
2.1.2. Límites de la generalización de una máquina de aprendizaje	60
2.1.3. Máquinas lineales de vectores soporte	62
2.1.4. Máquinas de vectores soporte lineales para regresión (SVR)	67
2.1.5. Multirregresor SVM	70
2.2. Redes de Procesos Gaussianos	71
2.2.1. Introducción	71
2.2.2. Regresión	72
2.2.2.1.1. El Modelo Lineal Estándar	72
2.2.3. Cálculo de la matriz de covarianzas	76
2.2.4. Sumario	77

3. Descripción de los algoritmos, datos y resultados	79
3.1. Descripción de los algoritmos	79
3.1.1. Modelo de datos	79
3.1.2. Modelo de señal	79
3.1.3. Modelo del algoritmo univariante: GLM	80
3.1.4. Modelo de los algoritmos multivariantes: GP y SVM	80
3.1.5. Descripción del problema	81
3.2. Datos	83
3.3. Sumario	83
3.4. Resultados	86
3.4.1. Caso 1: Matriz de covarianzas lineal	87
3.4.2. Caso 2: Matriz de covarianzas no lineal	100
4. Conclusiones	117
4.1. Discusión	117
4.2. Trabajos futuros	118
A. PRESUPUESTO DEL PROYECTO	119
B. Máquinas de vectores soporte no lineales	123
C. Proyecciones de entradas en el espacio de características	127
Bibliografía	129

ÍNDICE DE FIGURAS

1.1. Espín nuclear del protón del átomo de hidrógeno	23
1.2. Momento magnético. Magnetismo nuclear	24
1.3. Selección de corte. Gradiente y frecuencias de precesión asociadas.	25
1.4. Contraste según el instante de adquisición. Imágenes ponderadas. A) Imágenes en densidad protónica: poco contraste. B) Imágenes potenciadas en T1: son más anatómicas, más estructurales. C) Imágenes potenciadas en T2: son más fisiopatológicas y más funcionales.	27
1.5. Contraste según el instante de adquisición. Momentos longitudinal y transversal.	28
1.6. Eco y pulso de refase de la secuencia espín-eco.	30
1.7. Formación del espacio K. Secuencia de adquisición espín-eco.	31
1.8. Transformada de Fourier.	32
1.9. Tiempo de repetición y tiempo de eco.	33
1.10. Ejemplo de secuencia de eco de gradiente.	34
1.11. Mecanismo del efecto BOLD.	38
1.12. Evolución en el tiempo de los cambios neuronales, hemodinámicos y de señal de resonancia magnética.	39
1.13. Diseños paramétricos. Influencia de los cambios en la intensidad de una fuente luminosa sobre la actividad en la corteza visual. A medida que aumenta la intensidad del estímulo (de intensidad 3 a 6), se producen cambios en el patrón de actividad en corteza visual [15].	44
1.14. Diseño de bloques.	45
1.15. Tipos de diseño ligados a eventos.	48
1.16. Comparativa de calidad de imagen entre espín-eco y eco de gradiente para los mismos cortes.	51
1.17. Estudio de la inhomogeneidad y distorsión en la secuencia EPI mediante el paquete SPM. A) Imagen EPI eco de gradiente real. B) Distorsión a lo largo de diferentes direcciones.	52
1.18. Ajuste de la rejilla de Talairach a imagen anatómica.	54

1.19. Ajuste anatómico a cerebro normalizado MNI-EPI. A) Corte coronal y axial del cerebro MNI-EPI. B) Cortes coronal y axial de adquisición anatómica ajustada al cerebro normalizado MNI-EPI.	55
1.20. Promedio de épocas de activación y épocas de reposo para su posterior test estadístico. Comparación de actividad y reposo (A). Resultado final del estadístico para una $p < 0,006$ (B). Serie temporal promedio de un conjunto de píxeles activados (C).	56
2.1. El procedimiento SVM consiste en poner el hiperplano tan lejos como sea posible de las muestras más cercanas.	63
2.2. Hiperplanos de separación para el caso no separable.	66
2.3. Únicamente las muestras que estén fuera del margen $\pm \epsilon$, serán parte de la solución	68
2.4. Vapnik o función de coste ϵ -insensible.	68
3.1. Áreas del cerebro.	81
3.2. Corteza cerebral humana.	82
3.3. Diagrama de bloques del algoritmo univariante (GLM).	84
3.4. Diagrama de bloques de los algoritmos multivariantes (GP y SVM). . . .	85
3.5. Mapa de activación de β usando GLM del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	87
3.6. Mapa de activación del cociente entre el vector de pesos y la varianza usando GLM del estudio 1. A la derecha la rodaja con mayor activación cerebral.	87
3.7. Mapa de activación de β usando GP del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	88
3.8. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 1. A la derecha la rodaja con mayor activación cerebral.	88
3.9. Mapa de activación de β usando SVM del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	89
3.10. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 1. A la derecha la rodaja con mayor activación cerebral.	89
3.11. Mapa de activación de β usando GLM del estudio2. A la derecha la rodaja más significativa en la tarea motora.	90
3.12. Mapa de activación del cociente entre el vector de pesos y la varianza usando GLM del estudio 2. A la derecha la rodaja con mayor activación cerebral.	90
3.13. Mapa de activación de β usando GP del estudio 2. A la derecha la rodaja más significativa en la tarea motora.	91
3.14. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 2. A la derecha la rodaja con mayor activación cerebral.	91

3.15. Mapa de activación de β usando SVM del estudio 2. A la derecha la rodaja más significativa en la tarea motora.	92
3.16. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 2. A la derecha la rodaja con mayor activación cerebral.	92
3.17. Mapa de activación de β usando GLM del estudio 3. A la derecha la rodaja más significativa en la tarea motora.	93
3.18. Mapa de activación del cociente entre el vector de pesos y la varianza usando GLM del estudio 3. A la derecha la rodaja con mayor activación cerebral.	93
3.19. Mapa de activación de β usando GP del estudio 3. A la derecha la rodaja más significativa en la tarea motora.	94
3.20. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 3. A la derecha la rodaja con mayor activación cerebral.	94
3.21. Mapa de activación de β usando SVM del estudio 3. A la derecha la rodaja más significativa en la tarea motora.	95
3.22. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 3. A la derecha la rodaja con mayor activación cerebral.	95
3.23. Mapa de activación de β usando GLM del estudio 4. A la derecha la rodaja más significativa en la tarea motora.	96
3.24. Mapa de activación del cociente entre el vector de pesos y la varianza usando GLM del estudio 4. A la derecha la rodaja con mayor activación cerebral.	96
3.25. Mapa de activación de β usando GP del estudio 4. A la derecha la rodaja más significativa en la tarea motora.	97
3.26. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 4. A la derecha la rodaja con mayor activación cerebral.	97
3.27. Mapa de activación de β usando SVM del estudio 4. A la derecha la rodaja más significativa en la tarea motora.	98
3.28. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 4. A la derecha la rodaja con mayor activación cerebral.	98
3.29. Mapa de activación de β usando GP del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	100
3.30. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 1. A la derecha la rodaja con mayor activación cerebral.	100
3.31. Mapa de activación de β usando SVM del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	101

3.32. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 1. A la derecha la rodaja con mayor activación cerebral.	101
3.33. Mapa de activación de β usando GP del estudio 2. A la derecha la rodaja más significativa en la tarea motora.	102
3.34. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 2. A la derecha la rodaja con mayor activación cerebral.	102
3.35. Mapa de activación de β usando SVM del estudio 2. A la derecha la rodaja más significativa en la tarea motora.	103
3.36. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 2. A la derecha la rodaja con mayor activación cerebral.	103
3.37. Mapa de activación de β usando GP del estudio 3. A la derecha la rodaja más significativa en la tarea motora.	104
3.38. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 3. A la derecha la rodaja con mayor activación cerebral.	104
3.39. Mapa de activación de β usando SVM del estudio 3. A la derecha la rodaja más significativa en la tarea motora.	105
3.40. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 3. A la derecha la rodaja con mayor activación cerebral.	105
3.41. Mapa de activación de β usando GP del estudio 4. A la derecha la rodaja más significativa en la tarea motora.	106
3.42. Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 4. A la derecha la rodaja con mayor activación cerebral.	106
3.43. Mapa de activación de β usando SVM del estudio 4. A la derecha la rodaja más significativa en la tarea motora.	107
3.44. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 4. A la derecha la rodaja con mayor activación cerebral.	107
3.45. Mapa de activación de β usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	109
3.46. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	109
3.47. Mapa de activación de β usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	110
3.48. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	110

3.49. Mapa de activación de β usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	111
3.50. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	111
3.51. Mapa de activación de β usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	112
3.52. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	112
3.53. Mapa de activación de β usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	113
3.54. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	113
3.55. Mapa de activación de β usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	114
3.56. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	114
3.57. Mapa de activación de β usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	115
3.58. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	115
3.59. Mapa de activación de β usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.	116
3.60. Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.	116

ÍNDICE DE TABLAS

1.1. Valores T1 y T2 aproximados en cerebro.	29
A.1. Fases del Proyecto	120
A.2. Costes de personal	120
A.3. Costes de material	120
A.4. Costes de material	120
A.5. Presupuesto	121

RESONANCIA MAGNÉTICA FUNCIONAL

1.1. Introducción

La resonancia magnética (MR) es actualmente la técnica de imagen más utilizada en neurociencias, especialmente en lo referente a estudios estructurales. Dentro de las aplicaciones desarrolladas recientemente, la Resonancia Magnética funcional (fMR) constituye una revolución en el estudio de la actividad cerebral, no sólo por su resolución espacial, muy superior a la de las técnicas funcionales precedentes, y por la facilidad y exactitud para fusionar las imágenes funcionales con las estructurales, sino porque se trata de una técnica inocua, característica fundamental cuando se considera como método de investigación.

Si se considera el número de trabajos publicados sobre función cerebral en los últimos años, la fMR ocupa sin ninguna duda el primer lugar. Pero no todo son ventajas. La fMR es una técnica especialmente difícil: requiere un aparataje de alta calidad y por tanto de muy elevado coste; los métodos de adquisición y el diseño de los experimentos son muy complejos asimismo, el análisis de los datos requiere conocimientos profundos de física, informática y estadística.

1.2. Principios físicos de la resonancia magnética

MECANISMOS DE OBTENCIÓN DE LA SEÑAL DE RESONANCIA

Con diferencia, la técnica de obtención de imágenes por Resonancia Magnética es la que mayor dificultad conceptual conlleva de entre todas las modalidades de imagen médica. Le siguen muy de cerca, en cuanto a dificultad, la tomografía por emisión de positrones (PET) y la tomografía computarizada (TC). En primer lugar, porque una descripción rigurosa del fenómeno requeriría el uso de la mecánica cuántica, de gran dificultad incluso para los propios físicos. En segundo lugar, porque el tratamiento de la señal

obtenida en el escáner de resonancia requiere un seguimiento matemático informatizado.

Para que podamos obtener una imagen del interior de un objeto, sin que esto conlleve la rotura o descomposición de su estructura interna, es necesario hacer llegar hasta allí ondas con forma conocida. Ya sea por reflejo (ultrasonidos), por absorción y transmisión (TC), o por emisión desde el interior (PET y MR), e intentar captar la modificación que se ha producido en esa onda ya sea en su fase, en su frecuencia o en su amplitud.

El tipo de ondas que se emplean en resonancia para bombardear los tejidos son ondas electromagnéticas a frecuencias de radio del orden de Megahertzios. Los receptores, y posteriormente los emisores, del interior del cuerpo humano son los protones de algunos núcleos atómicos, que hacen de antena emisora y receptora. Los protones implicados en la imagen de resonancia, ya sea imagen anatómica o funcional, son los protones de los átomos de hidrógeno. En el cuerpo humano hay gran cantidad de átomos de hidrógeno, sobre todo por la cantidad de moléculas de agua que hay en el organismo. El núcleo del átomo de hidrógeno sólo contiene un único protón, el cual tiene una propiedad denominada espín y que le hace comportarse como un pequeño imán con su polo norte y sur. Este espín también se manifiesta, siempre desde la física clásica, como una rotación sobre su eje (fig.1.1). Dependiendo del sentido de la rotación del protón decimos que está en estado paralelo o en estado antiparalelo y se representa con una flecha que apunta, o bien hacia arriba (estado paralelo), o bien hacia abajo (estado antiparalelo). Arriba, abajo, paralelo o antiparalelo, son asignaciones arbitrarias con objeto de tener una idea intuitiva de los dos posibles estados que puede mantener un protón cuando interacciona con campos magnéticos.

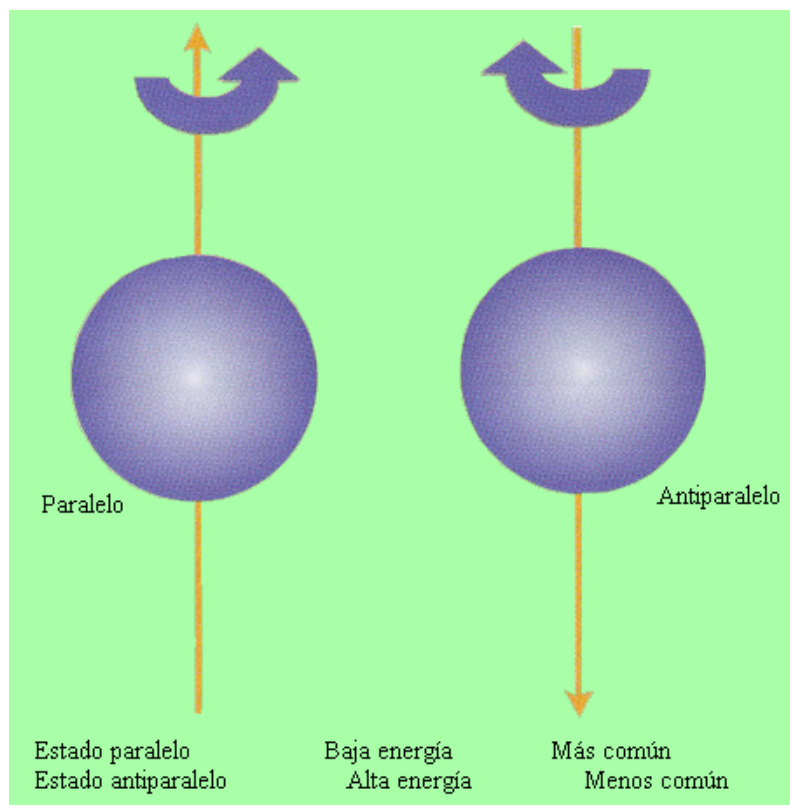


Figura 1.1: Espín nuclear del protón del átomo de hidrógeno

En un solo gramo de agua puede haber del orden de 10^{23} protones de este tipo. Esta cantidad tan elevada hace que sea posible detectar con antenas la señal electromagnética que pueden llegar a emitir si les aportamos una energía extra y también es la razón de que en realidad lo que captemos sea el comportamiento promedio de un puñado de ellos (fig.1.2).

En la figura se ha representado esquemáticamente lo que sería una imagen, formada por elementos cúbicos que serían sus píxeles (en realidad son vóxeles o elementos de volumen). Si extraemos un vóxel de la imagen, su nivel de gris corresponde a millones de billones de protones. El comportamiento magnético promedio de todos ellos viene descrito por un vector denominado momento magnético. No todos los protones de ese vóxel van a emitir sólo una fracción de aquellos que se encuentran inicialmente en el estado de baja energía o estado paralelo, ya que estos son los únicos que pueden adquirir energía para posteriormente emitirla en forma de ondas electromagnéticas. De cara a la intensidad de señal que recoge el escáner, cuantos más protones haya en estado de baja energía, mejor, pues potencialmente habría una mayor población de protones susceptibles a la absorción y posterior emisión de energía. Desgraciadamente, esta distribución de protones en un estado u otro depende de la temperatura a la que se encuentre la muestra y está perfectamente descrito por la distribución probabilística de Boltzmann.

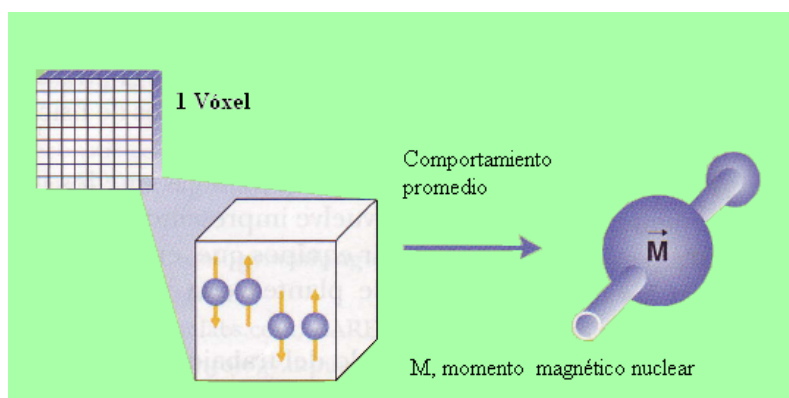


Figura 1.2: Momento magnético. Magnetismo nuclear

Puesto que la temperatura del cuerpo humano es de 36-37°C, muchos de los protones se encuentran en el estado de mayor energía. El escáner de resonancia consta de un campo magnético muy intenso, homogéneo y uniforme. Habitualmente tiene intensidades de 0,5 a 4T para humanos, pasando por 1T, 1,5T y 3T. Cuando situamos al organismo bajo la influencia de ese campo, los protones de los tejidos se alinean con respecto a él. Este alineamiento puede ser en sentido paralelo o antiparalelo. El proceso de alineamiento va acompañado de un movimiento de precisión similar al de una peonza, es decir, rota en torno a su propio eje pero también gira en torno a la dirección del eje del campo magnético hasta que consigue alinearse con él. Este movimiento de precisión lo realiza a una velocidad angular, denominada frecuencia de precesión de Larmor, cuyo valor depende, únicamente, del producto de dos parámetros: la constante giromagnética del protón del núcleo del hidrógeno cuyo valor es 42,6 MHz por cada tesla, y la intensidad de campo magnético. Es decir, que si el campo magnético del escáner de resonancia fuera de 1T, la frecuencia de Larmor sería:

$$w = 42,6 \frac{\text{MHz}}{\text{tesla}} \times 1\text{T} = 42,6\text{MHz} \quad (1.1)$$

Por lo tanto, conseguiríamos que todos los protones del hidrógeno del interior del cuerpo humano precesasen a una frecuencia de 42,6MHz (si el campo magnético fuera de 1T).

Si queremos mandar energía a los protones que se encuentran en el estado de baja energía, lo podríamos hacer de dos formas: o bien, elevando la temperatura corporal, o bien, enviando una onda de radio a la frecuencia de precesión de los protones o frecuencia de Larmor. En nuestro ejemplo, una onda de radio de 42,6 MHz sería absorbida por una cantidad de protones en estado de baja energía que, dependiendo de la intensidad y duración del pulso, les permitiría pasar del estado de baja a alta energía (del estado paralelo al estado antiparalelo). El hecho de que algunos de ellos comiencen a cambiar de estado afecta al momento magnético nuclear. Éste, por ser un valor vectorial que promedia una población de protones, tiene un comportamiento análogo a uno de ellos. También precesa entorno a la dirección principal del campo magnético y puede variar su orientación de

forma continua a medida que los protones pasan de un estado a otro. La cantidad de protones que absorben energía, depende de la duración e intensidad de la energía recibida por la muestra. Esta energía proviene del pulso de radio a la frecuencia de precesión. El comportamiento matemático del vector momento magnético está perfectamente descrito mediante las denominadas ecuaciones de Bloch. Una acertada combinación de pulsos de radio, denominada secuencias de adquisición, permite gobernar el movimiento del momento magnético, haciendo que se desplace un determinado número de grados respecto del eje del campo magnético principal.

Para seleccionar un corte perpendicular al campo magnético principal, es decir, un corte axial se introduce un gradiente de campo magnético a lo largo del eje del campo magnético principal. Un gradiente es una variación lineal de la intensidad del campo magnético que depende de la posición. Es decir, lo que se hace es variar levemente la intensidad de campo magnético que experimentan los protones a lo largo de dicho eje (fig.1.3). Al experimentar diferentes intensidades de campo magnético también precesan a distintas frecuencias de Larmor y, por tanto, ya no absorben la energía a la misma frecuencia. La forma de seleccionar el corte es enviar un pulso a la frecuencia de precesión de los protones de la región de interés en presencia de un gradiente. Este pulso se denomina de excitación o pulso de selección de corte.

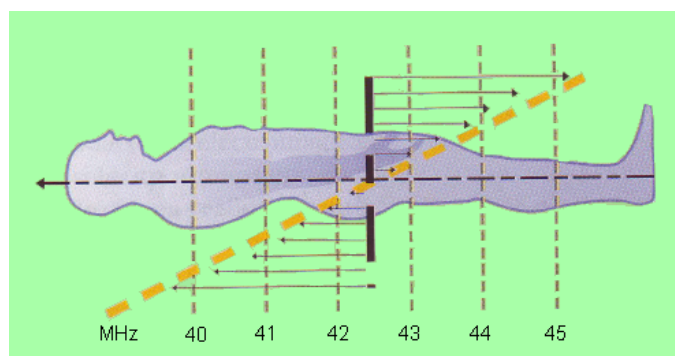


Figura 1.3: Selección de corte. Gradiente y frecuencias de precesión asociadas.

Técnicamente es imposible lanzar un pulso a una frecuencia exacta. Lo que en realidad se envía es un pulso centrado en una frecuencia pero con una cierta inexactitud o incertidumbre en anchura. La anchura en frecuencia del pulso nos permite afectar a protones que precesan a frecuencias próximas a la frecuencia de precesión del centro de corte seleccionado. Cuanta más anchura en frecuencia, más anchura de corte. Cuanta más anchura de corte, más señal, es decir, mejor relación señal a ruido, pues habrá más cantidad de protones resonando y, por lo tanto, emitiendo. Sin embargo, se pierde en resolución espacial, al menos en esa dimensión, a lo largo del eje del campo magnético principal.

Una vez que cesa el pulso de radiofrecuencia, los protones que han absorbido energía pasando al estado antiparalelo tienden a volver al estado de baja energía o paralelo (a este proceso se le denomina *relajación*) y esto lo hacen, emitiendo una señal de radio a la fre-

cuencia de precesión. Esta señal recibe el nombre de “señal de decaimiento de inducción libre” (*free induction decay*, **FID**). Esa señal contiene información de todo el corte.

CONTRASTE EN RESONANCIA MAGNÉTICA. TIEMPOS DE RELAJACIÓN T1 Y T2

El contraste de la imagen viene dado por los diferentes niveles de gris. Esto nos permite distinguir tejidos. El contraste es un fenómeno complejo que depende de diversos factores: la densidad de protones de un tejido, los tiempos de relajación T1 y T2, y el instante en que tomemos la imagen.

Dependiendo de la posición en el tiempo de ese instante, tendremos, o bien una *imagen en densidad protónica*, o bien una *imagen ponderada en T1*, o bien una *imagen ponderada en T2*. Esto afecta a la intensidad de la señal dependiendo del tipo de tejido. Las imágenes de densidad protónica son las de más escaso contraste, ya que este depende, únicamente, de la cantidad absoluta de protones del hidrógeno en cada tejido. Como el agua abunda en el cuerpo humano, la cantidad de hidrógeno es abundante en casi todos los tejidos, salvo en el tejido óseo, por lo que la intensidad de señal procedente de cada tejido va a ser muy similar (fig.1.4). Sin embargo, las imágenes ponderadas, o bien en T1, o bien en T2, van a ser imágenes con muy alto contraste. Las imágenes ponderadas en T1 están más orientadas a la anatomía permitiendo detectar patologías que cursen con cambios morfológicos. Las imágenes ponderadas en T2 están más relacionadas con la fisiopatología. Por esto último, la imagen funcional es imagen ponderada en T2. El contraste depende, intrínsecamente, de la diferente manera en la que se produce la relajación de los protones del átomo de hidrógeno dependiendo del entorno fisicoquímico en que se encuentre. Es decir, va a depender de la molécula y tejido donde se encuentre dicho protón. En la figura 1.4 podemos observar las diferencias sustanciales entre los tipos de imagen en el caso del cerebro. Observamos cómo, obviamente, la imagen en densidad protónica está peor contrastada frente a la imagen en T1 o la imagen en T2. También son muy apreciables las diferentes maneras de brillar de un tejido ya sea en T1 o en T2.

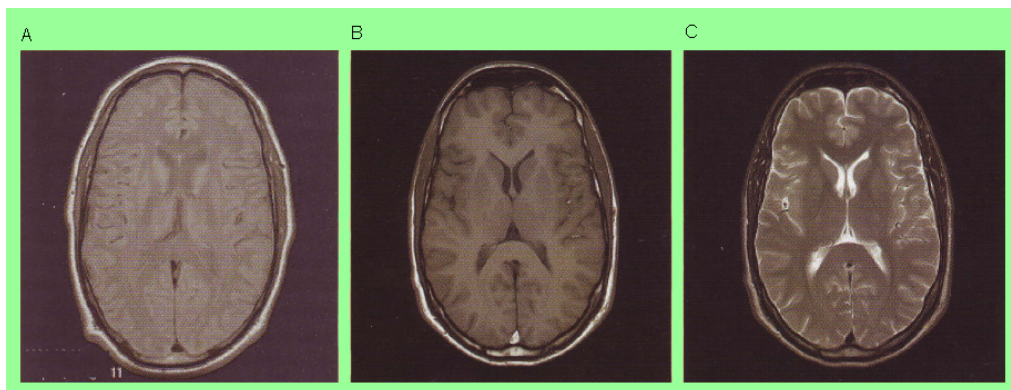


Figura 1.4: Contraste según el instante de adquisición. Imágenes ponderadas. A) Imágenes en densidad protónica: poco contraste. B) Imágenes potenciadas en T1: son más anatómicas, más estructurales. C) Imágenes potenciadas en T2: son más fisiopatológicas y más funcionales.

Cuando lanzamos el pulso de excitación (se hace girar el momento magnético) un cierto ángulo dentro del plano formado por los ejes X y Z (generalmente 90° en la secuencia más habitual, denominada secuencia *espín-eco*). Una vez que cesa el pulso de radiofrecuencia, los protones comienzan a regresar al estado original de baja energía. También el momento magnético comienza a regresar a la posición original. El tiempo T1, por tanto, está relacionado con la duración de la relajación hasta recuperar la posición original. Si observamos la figura 1.5, podemos hacernos una idea de este movimiento en espiral de un eje imaginario solidario con el vector momento magnético. En la misma figura podemos ver dibujadas dos proyecciones del vector momento magnético:

- *proyección longitudinal o magnetización longitudinal*: sobre el eje z.
- *proyección transversal o magnetización transversal*: sobre el plano xy o plano de corte.

Ambas son, las componentes longitudinal y transversal de la magnetización. A medida que transcurre el tiempo, la muestra irá ganando en magnetización longitudinal (pasando de valor cero, cuando el vector momento magnético está sobre el plano de corte, a valor máximo cuando se encuentre sobre el eje z) y perdiendo magnetización transversal (pasando del valor máximo, cuando se encuentre sobre el plano xy, a valor cero, cuando el vector momento haya regresado a su posición original sobre el eje z).

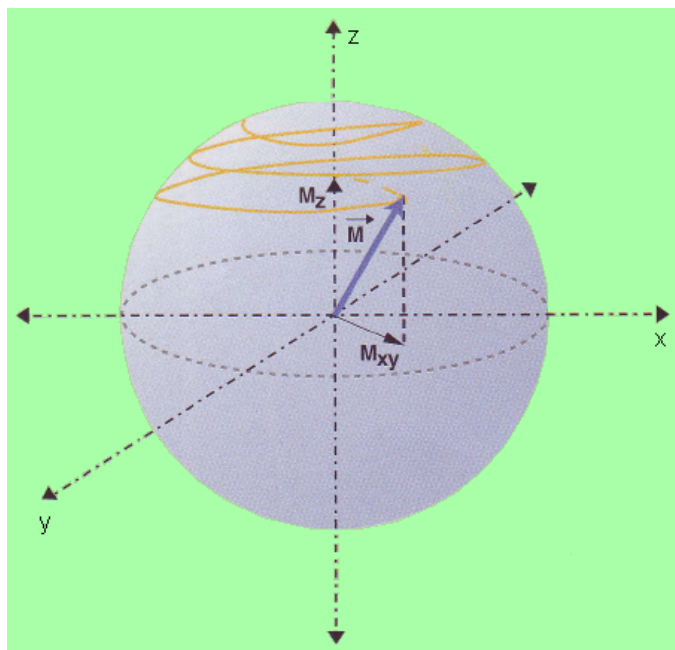


Figura 1.5: *Contraste según el instante de adquisición. Momentos longitudinal y transversal.*

Podríamos pensar que se gana magnetización longitudinal al mismo ritmo que se pierde magnetización transversal. Pues bien, no ocurre así. Si imaginamos las proyecciones sobre el eje z y sobre el plano xy como sendas sombras que proyecta el vector momento magnético, curiosamente, antes de que se haya recuperado completamente la magnetización longitudinal, la sombra sobre el plano xy ha desaparecido. Es decir, desaparece la magnetización transversal antes de que se haya recuperado la magnetización longitudinal. Por este motivo podemos considerar un tiempo $T1$, vinculado al proceso de recuperación de la magnetización longitudinal de la muestra, y un tiempo $T2$ vinculado a la pérdida de magnetización transversal.

En concreto, podemos definir:

- *Tiempo $T1$* : el tiempo que tarda en recuperar un 63 % de la magnetización longitudinal.
- *Tiempo $T2$* : el tiempo que tarda en perder un 63 % de la magnetización transversal.

La razón de que se pierda la magnetización transversal más rápidamente de lo que se recupera la longitudinal tiene su explicación en cómo se construye el vector momento magnético de la muestra y en la interacción de los protones del átomo de hidrógeno con su entorno molecular. Debido a la polaridad de ciertas macromoléculas se generan inhomogeneidades microscópicas que hacen que no todos los protones incluidos en el mismo vóxel experimenten la misma intensidad de campo magnético. Esto provoca que en sus movimientos de precesión, unos protones se adelanten (o retrasen) con respecto a otros,

es decir, se desfasan. Este fenómeno es conocido como *decoherencia*.

El T1 también se denomina "tiempo de relajación *spin-lattice*", por estar relacionado con la capacidad del tejido para absorber en forma de aumento de temperatura la radiación que emiten los protones. El T2 también se denomina "tiempo de relajación *spin-spin*", por estar más relacionado con la interacción que tiene el espín de un protón con los espines de los protones que le circundan.

En la siguiente tabla aparecen unos valores aproximados de T1 y T2 para las principales sustancias y tejidos del cerebro.

	T1(MS)	T2(MS)
Materia blanca	510	67
Materia gris	760	77
Edema	900	126
Líquido cefalorraquídeo	2650	180

Tabla 1.1: Valores T1 y T2 aproximados en cerebro.

En la práctica, el tiempo T2 siempre es más corto que el tiempo T1 pero, además, es más breve de lo que las ecuaciones de Bloch describen y predicen. Esto es debido a que a las inhomogeneidades microscópicas hay que añadir otro elemento que contribuye al desfase, que es la inhomogeneidad de campo magnético externo. Es imposible construir un campo magnético tan extenso e intenso y que sea perfectamente homogéneo en todos los puntos de la muestra. Estas pequeñas inhomogeneidades de campo, unidas a las inhomogeneidades generadas por las macromoléculas provocan la aparición de un tiempo más breve aún que el T2 y que se denomina T2 asterisco (T2*).

El que haya inhomogeneidades debidas a la acción de ciertas macromoléculas no deja de ser una ventaja que nos permite establecer contrastes entre tejidos. Sin embargo, las inhomogeneidades debidas al campo magnético externo son muy negativas de cara a la obtención de la imagen pues acortan el tiempo de adquisición, por lo que tendremos peor relación señal ruido, y pueden aparecer diferencias de gris locales no debidas a diferencias en la composición sino a la inhomogeneidad de campo. Sin embargo, es fácil cambiar el T2* por el T2, sin más que suministrar a la muestra un pulso de refase (lo que se denomina *secuencia Hahn spin-eco*). Consiste en que, una vez que comienza el desfase, se envía un pulso de radio, centrado también en la frecuencia de precesión, que hace voltear 180° los espines protónicos (figura 1.6). Los que se retrasaban son colocados por delante, y los que se adelantaban se colocan por detrás por lo que, al cabo de un cierto tiempo, los lentos son alcanzados por los rápidos volviendo a juntarse (refase) durante un cierto tiempo, para luego, evidentemente, volver a desfasearse. La ventaja es que se ha conseguido compensar, brevemente, el efecto de desfase. Si lanzamos un pulso cuya intensidad y duración provoque un volteo de 180°, los protones pueden refasar formándose lo que

se denomina un eco. De este modo, hemos pasado de $T2^*$ a tener un tiempo $T2$. Es precisamente ese eco el que vamos a recoger mediante una antena. Ese eco muestreado es la información previa para la posterior formación de la imagen.

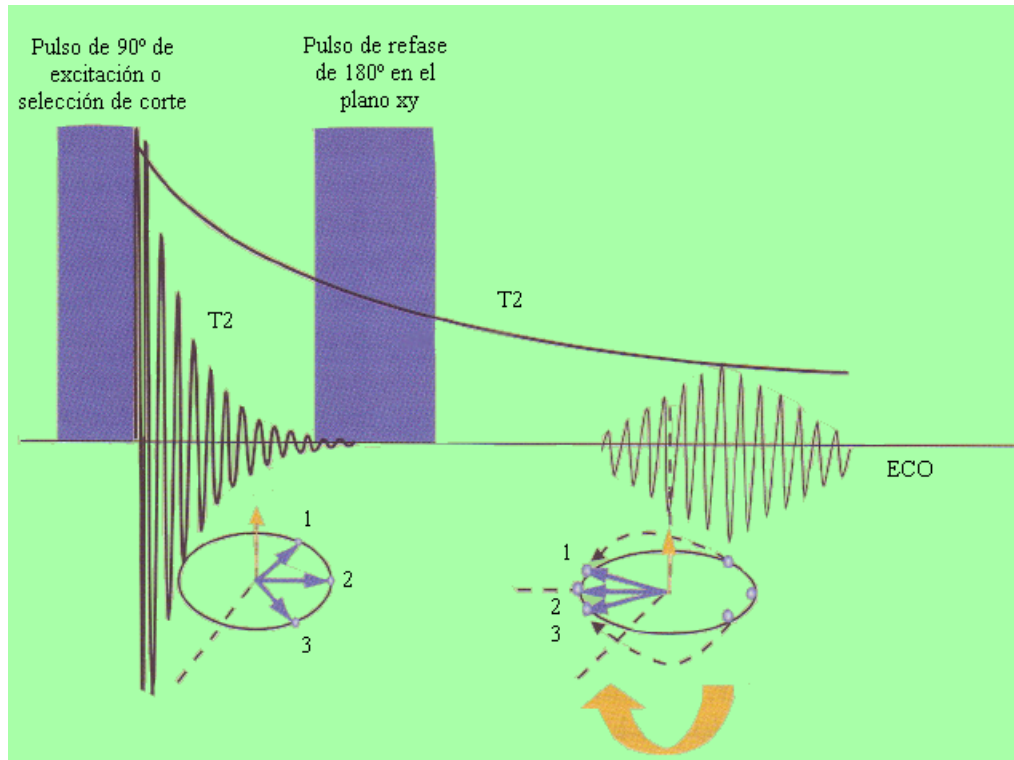


Figura 1.6: Eco y pulso de refase de la secuencia espín-eco.

TÉCNICAS DE IMAGEN EN RESONANCIA MAGNÉTICA. FORMACIÓN DE LA IMAGEN. SECUENCIAS. TIEMPO DE ECO Y TIEMPO DE REPETICIÓN. EL ESPACIO K

La formación de la imagen de resonancia a partir de los ecos requiere de un procedimiento complejo y de una descripción, también compleja de la imagen. El procedimiento son las denominadas *secuencias de adquisición*. Por otra parte, una descripción muy particular de la imagen es lo que se denomina *espacio K*. En la figura 1.7 se muestra la secuencia espín eco. En realidad, una secuencia de adquisición es el conjunto de pulsos y gradientes necesarios para poder adquirir el número de ecos necesarios para formar el espacio K. Los escáneres de resonancia lo que realmente adquieren no es la imagen en sí, sino el espacio K que contiene las frecuencias espaciales de la imagen buscada. La imagen final se reconstruye en un ordenador mediante una sofisticada operación matemática denominada *transformada de Fourier* del espacio K obtenido en el escáner. El espacio K está formado por las muestras tomadas de los ecos generados de forma sucesiva.

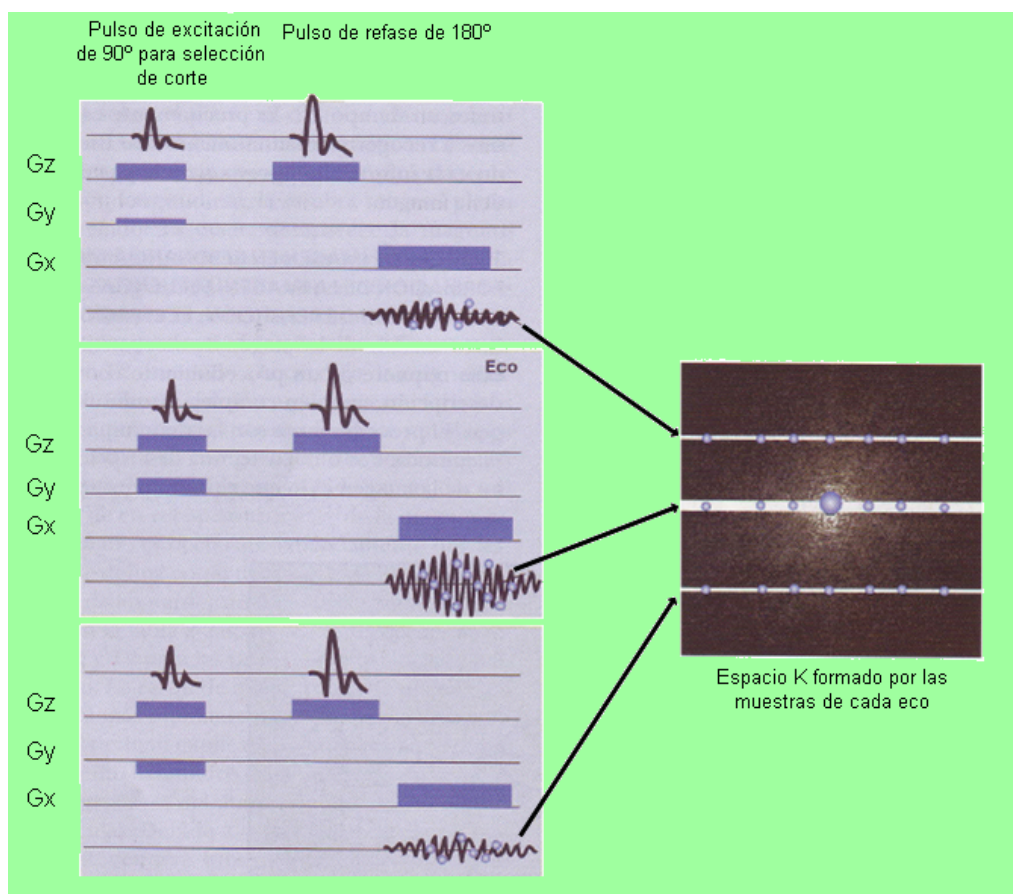


Figura 1.7: Formación del espacio K. Secuencia de adquisición espín-eco.

En resumen, la descripción de una secuencia es: introducir un gradiente en el eje z , a continuación un pulso de radio centrado en la frecuencia de los protones del hidrógeno del corte de interés (pulso de excitación), a continuación pulso de refase de 180° y, por último, la adquisición propiamente dicha, mediante muestreo del eco de la señal en una antena. Cada eco obtenido para generar cada línea del espacio K está generado de forma diferente. En concreto variando dos gradientes perpendiculares al eje z . El gradiente a lo largo del eje y , g_y se denomina *gradiente de codificación de fase* y es modificado para obtener cada línea del espacio K . El gradiente a lo largo del eje x , g_x es el eje de codificación de frecuencia. También se le denomina gradiente de lectura, pues, a diferencia del gradiente de fase, suele estar activo mientras se realiza el muestreo de la señal. En la secuencia espín-eco, para generar cada eco se necesita un pulso de excitación y variar el gradiente de codificación de fase. Si la imagen tiene 256 filas de píxeles, necesitaremos 256 ecos. Es decir, tantos ecos como filas en la imagen. Además, necesitamos obtener tantas muestras del eco como columnas de píxeles necesitemos. Dicho de otro modo, las dimensiones de la imagen coinciden exactamente con las dimensiones del espacio K . En la figura 1.8 se puede ver el espacio K y la imagen obtenida tras realizar en un ordenador la transformada inversa de Fourier.

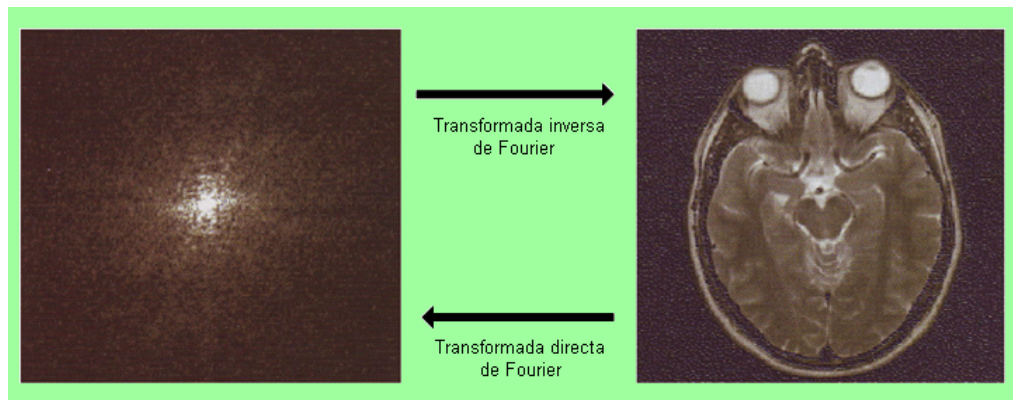


Figura 1.8: Transformada de Fourier.

A continuación introducimos dos conceptos relacionados con el T_1 y T_2 que son el tiempo de repetición (TR) y el tiempo de eco (TE). El **TR** es el tiempo que transcurre entre dos pulsos de excitación sucesivos para obtener sendas líneas del espacio K . El **TE** es el tiempo que transcurre entre el pulso de excitación y la aparición del eco. Ajustando estos tiempos es como podemos ponderar en T_1 y en T_2 u obtener imágenes en densidad protónica. En la figura ?? podemos observar la relación entre los TE y TR y la ponderación en T_1 y T_2 , y la imagen de densidad protónica.

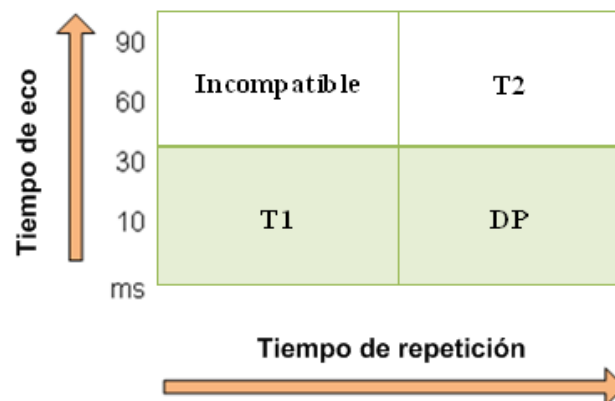


Figura 1.9: Tiempo de repetición y tiempo de eco.

Hay tres tipos de secuencias básicas:

- *espín-eco*: son las más comunes.
- *inversión-recuperación*: son interesantes porque permiten resaltar tejidos de forma selectiva. Se basan en suministrar pulsos de excitación de 180° en lugar de 90° . Esto provoca que cambie la intensidad de gris e incluso se pueda suprimir algún tejido que dificulte el diagnóstico sobre la imagen.
- *eco de gradiente*: son empleadas en fMR pues permiten adquirir las imágenes mucho más rápidamente y así poder realizar el seguimiento de la fisiología.

SECUENCIAS RÁPIDAS

La forma de obtener rápidamente una imagen de resonancia tiene mucho que ver con la manera en que se construye el espacio K. El espacio K es un espacio hermítico. Esto significa que es simétrico respecto de la diagonal. Por tanto, una forma más rápida de conseguir una imagen es obteniendo el espacio K de forma parcial y, posteriormente, rellenar lo que falta aprovechando la simetría propia de un espacio hermítico. Esta técnica se denomina *Half-Fourier*. Otra posibilidad para acelerar la adquisición es dando pulsos de excitación que hagan que el momento magnético se incline un ángulo menor a 90° . Por último, las secuencias rápidas de adquisición empleadas en resonancia magnética funcional son las secuencias eco de gradiente, como, por ejemplo, la secuencia “imagen ecoplanar” o secuencia EPI (*echo-planar imaging*).

En la figura 1.10 podemos ver el diagrama de pulsos. Lo característica de esta secuencia es que puede llegar a obtener todos los ecos necesarios con un solo pulso de excitación o selección de corte. Los ecos se obtienen haciendo variar alguno de los gradientes (o bien el de base, o bien el de lectura) sin aplicar pulso de refase. El gradiente opera de manera que, primero fuerza un desfase para, a continuación, forzar un refase.

En la figura se ha representado una de las muchas posibilidades de secuencia de eco de gradiente mediante un gradiente de lectura negativo y otro positivo. El hecho de limitar el número de pulsos de excitación empeora la relación señal-ruido, a cambio de acelerar la adquisición. También provoca la aparición de artefactos, como el denominado *Ghosting*, que se caracteriza por la aparición de bordes de la imagen duplicados y desplazados, o la ausencia de señal en lugares próximos al hueso. Con las técnicas de eco de gradiente se pueden llegar a obtener imágenes en 30 ms.

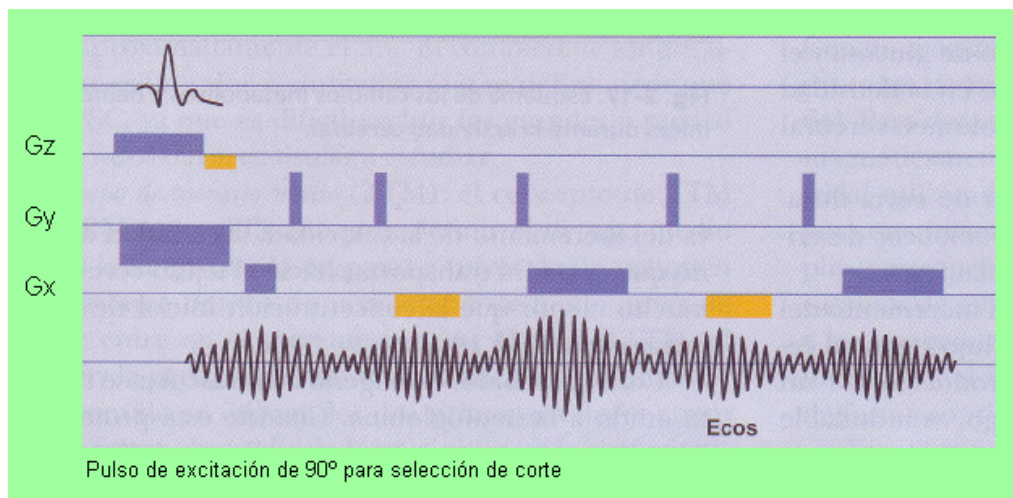


Figura 1.10: Ejemplo de secuencia de eco de gradiente.

1.3. Bases neurofisiológicas de la resonancia magnética funcional

METABOLISMO CEREBRAL Y FLUJO SANGUÍNEO REGIONAL

El funcionamiento cerebral se basa en un intercambio de estímulos eléctricos. Estas incitaciones eléctricas se producen en las células y existen varios niveles de activación cerebral. Las membranas celulares a través de las bombas de sodio y potasio generan unos potenciales eléctricos de membrana que delante de unos estímulos se convierten en potenciales de acción. Estos cambios son los que permiten la sinapsis que consiste en el intercambio de neurotransmisores entre neuronas.

Todos estos procesos que se ponen en marcha al realizar una acción cerebral consumen energía. Por lo tanto, debido a la constante actividad cerebral el requerimiento energético es continuo. El cerebro debe generar altas cantidades de adenosín trifosfato (ATP) como fuente de energía para mantener estos potenciales de membrana esenciales para la transmisión de impulsos nerviosos. En condiciones normales el cerebro utiliza

glucosa como único metabolito energético para satisfacer la gran demanda energética cerebral, consumiendo el 60 % de la glucosa total requerida por un individuo en reposo. Este órgano necesita alrededor de 120 g de glucosa por día, lo que corresponde al 15 % de la energía total consumida cada día. El cerebro también es un órgano altamente aerobio, y su metabolismo utiliza más del 20 % del oxígeno total consumido por un individuo. Es importante recordar que el cerebro no dispone de otros metabolitos de reserva como sería el glucógeno y, por lo tanto, el suministro de glucosa y oxígeno no puede ser interrumpido ya que se produciría un daño anóxico cerebral.

La glucosa, así como el oxígeno y otras sustancias, son enviados a las neuronas activas metabólicamente vía perfusión sanguínea a través de los lechos capilares, produciéndose un incremento del flujo regional cerebral. Así pues, los cambios en el flujo cerebral regional, los cambios en el volumen sanguíneo cerebral y el consumo de glucosa o de oxígeno son parámetros indirectos que reflejan la actividad cerebral [1].

Cuando se aplica un estímulo cerebral durante un período largo, inicialmente se detecta un incremento del consumo de glucosa y de oxígeno y existe una relación directamente proporcional entre estos dos parámetros. Se produce un aumento del flujo cerebral regional para compensar este aumento en el requerimiento de oxígeno. Si el estímulo es duradero, para mantener la constante entre el consumo de glucosa y el de oxígeno, se producirá un incremento en la densidad capilar y, por lo tanto, un aumento del volumen cerebral total [2],[3],[4]. Sin embargo, cuando el estímulo es de corta duración, los cambios hemodinámicos y metabólicos no se reproducen de igual forma.

Estudios con PET han demostrado el incremento del consumo de glucosa, de oxígeno y del flujo regional cerebral durante la activación cerebral producida por un estímulo de corta duración. Sin embargo, es indudable que en un período corto no se puede incrementar la densidad capilar, por lo que se produce un incremento de la velocidad del flujo sanguíneo provocado por cambios en el diámetro capilar.

El oxígeno intravascular se transporta unido a la hemoglobina. Cuando esta proteína está llena de oxígeno se denomina oxihemoglobina y cuando lo cede se transforma en deoxihemoglobina. Durante la activación cerebral se producirá un incremento local de oxihemoglobina arterial y capilar, sin embargo, la concentración de deoxihemoglobina disminuirá debido al descenso del transporte de oxígeno tisular. Esta caída de concentración de la deoxihemoglobina provocará un aumento de señal en las imágenes de fMR.

En resumen, los cambios metabólicos y hemodinámicos que acontecen durante la actividad cerebral son herramientas que nos permitirán estudiar el funcionamiento cerebral. Los estudios de PET se basan mayoritariamente en los cambios metabólicos de la glucosa, mientras que en resonancia magnética los estudios se asientan básicamente en los cambios hemodinámicos de perfusión y de flujo sanguíneo.

RESONANCIA MAGNÉTICA Y VOLUMEN SANGUÍNEO REGIONAL: VOLUMEN CEREBRAL SANGUÍNEO Y ARTERIAL SPIN LABELING

En los últimos tiempos se han desarrollado varias técnicas de resonancia magnética que permiten calcular el flujo sanguíneo cerebral (FSC). Los métodos más utilizados hasta ahora son el rastreo dinámico de un bolo de contraste paramagnético, denominado *imagen de perfusión* y el denominado *arterial spin labeling*.

Algunos de los parámetros que se deben conocer al realizar estudios de perfusión con resonancia magnética son:

- *Perfusión o flujo sanguíneo cerebral (FSC)*: es la velocidad con la que el flujo cerebral pasa a través de la microvasculatura cerebral. Es un concepto fisiológico independiente de los parámetros experimentales de la resonancia magnética. La perfusión se define como el volumen de sangre que pasa a través de una cierta masa de tejido en un período determinado. La unidad que se utiliza normalmente es mililitro por 100 mg/min.
- *volumen sanguíneo cerebral (VSF)*: es la fracción del volumen de tejido que es ocupado por sangre. El volumen medio de sangre en el cerebro es de aproximadamente el 3%.
- *Tiempo de tránsito medio (TTM)*: es el tiempo medio que la sangre tarda en pasar a través de un volumen sanguíneo dentro de una región de tejido antes de que entre en el torrente venoso.

En la actualidad, el método de resonancia magnética más utilizado para estudiar la perfusión cerebral desde el punto de vista clínico es la técnica de la inyección de contraste paramagnético en forma de bolo, también denominada *técnica de bolo tracking*. Esta técnica consiste en la inyección de un bolo de un agente de contraste y una medición rápida de la pérdida de señal que se produce provocada por el desfase de los espines. Este desfase se produce al existir más de un campo magnético en el mismo vóxel. Los vasos sanguíneos que llevan una alta concentración de un agente de contraste con efecto paramagnético tienen diferentes campos magnéticos que el tejido circundante, todo ello provocando una caída del $T2^*$ y a su vez una disminución de la señal en las secuencias potenciadas en $T2$. Como el tiempo de tránsito del bolo a través del tejido es muy rápido, es necesario disponer de una alta resolución temporal para obtener imágenes secuenciales durante la entrada y la salida del contraste paramagnético y de esta forma cuantificar el FSC y el VSF. El primer estudio que utilizó técnicas de perfusión con resonancia magnética para observar la actividad funcional en humanos fue realizado por JW Belliveau et al. [5], estableciendo un mapa del córtex visual al comparar dos estudios de perfusión realizados mediante la administración de contraste antes y durante una estimulación visual. Aunque su publicación levantó muchas expectativas, la búsqueda de nuevos contrastes endógenos como la técnica de BOLD o el ASL relegaron los estudios de fMR mediante perfusión con técnica de *bolo tracking* a un segundo plano.

La técnica de ASL cobra cada vez más auge. Esta técnica se basa en la marcación de los protones que circulan por el torrente sanguíneo. Es conocido el fenómeno de aflujo en resonancia magnética, que consiste en una ausencia de señal en el interior de los vasos sanguíneos provocada porque en su interior los hidrogeniones circulan a gran velocidad y no se saturan cuando el TR es corto.

Estos protones no saturados una vez localizados en los capilares prefunden hacia el parénquima provocando cambios de señal. Cuanto más grande sea el grado de perfusión o flujo de protones dentro del corte, mayor será el aumento de señal. Para determinar el flujo cerebral se marcan electromagnéticamente los hidrogeniones entrantes por vía arterial utilizando un pulso de radiofrecuencia diseñado para que invierta la dirección de los protones en un amplio corte próximo al plano a estudiar. Midiendo los cambios de señal obtenidos entre las imágenes marcadas y las imágenes control no marcadas se puede obtener imágenes del FSC.

Esta técnica de perfusión inicialmente utilizada para estudios de isquemia cerebral se está aplicando también para realizar estudios de actividad cerebral. La capacidad de esta técnica de poder cuantificar el FSC probablemente permita localizar mejor la actividad cerebral que las técnicas BOLD y ayude a la mayor comprensión fisiológica que ocurre durante la actividad cerebral entre diferentes regiones cerebrales.

Sin embargo, y a pesar de las grandes expectativas del ASL, esta técnica aún debe de mejorar, sobre todo en lo referente a la adquisición multicorte.

MECANISMO DEL EFECTO BLOOD OXYGEN LEVEL DEPENDENT

Sin duda la técnica de fMR más utilizada en la actualidad es la que emplea el efecto BOLD (*blood oxygen level dependent*). Esta técnica se basa en la susceptibilidad magnética de la hemoglobina que se comporta como un contraste endógeno que nos permitirá detectar cambios en el flujo sanguíneo cerebral.

Ogawa et al. fueron los primeros en detectar que cambios en la oxigenación sanguínea podían provocar alteraciones en el parámetro de la caída del T2* conduciendo a cambios en la intensidad de señal en las secuencias potenciadas en T2 [6]. Este efecto se debe a la susceptibilidad magnética de los diferentes estados de la Hemoglobina (Hb); la deoxihemoglobina es paramagnética, mientras que la Hb totalmente oxigenada (oxihemoglobina [oxiHb]) es diamagnética. La presencia de la deoxihemoglobina altera el campo magnético local y permite una caída más rápida de la señal acortando el T2*. Cuanta más deoxihemoglobina exista más rápida será la caída de la señal y menor será la señal detectada. A la inversa, más oxiHb conduce a una caída de la señal más lenta. El T2* local utilizado como un contraste de resonancia magnética se determina por la proporción entre deoxihemoglobina y oxihemoglobina dentro de un área particular del cerebro. Un aumento de flujo sanguíneo con alta concentración de sangre oxigenada en una región de interés particular, provocado por una activación neural, conducirá a un au-

mento relativo del $T2^*$ en esa región. Eso a su vez se traducirá en un incremento de la intensidad de la señal. Así pues, la deoxihemoglobina es un agente de contraste endógeno y sirve como fuente de señal para aplicaciones de fMR [7] (fig. 1.11).

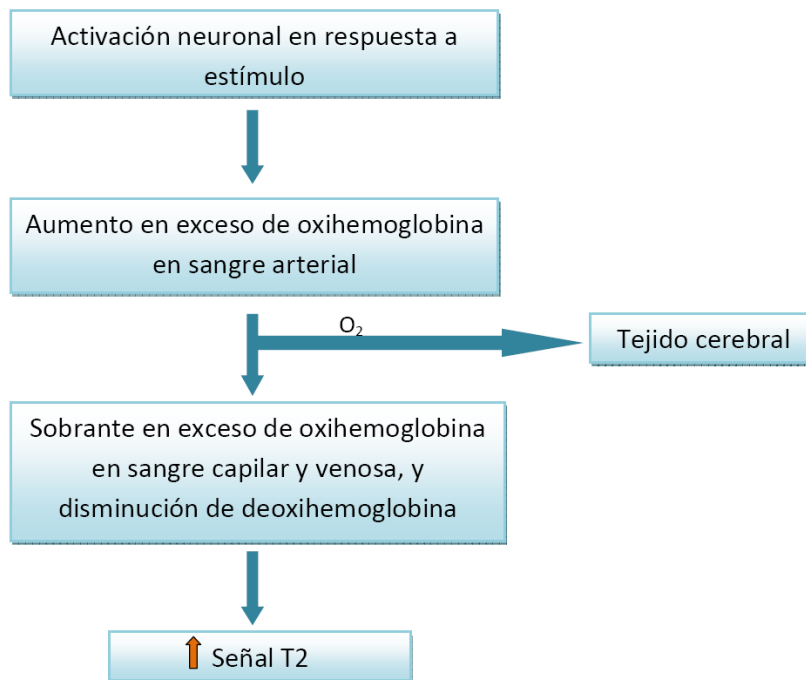


Figura 1.11: Mecanismo del efecto BOLD.

Inmediatamente después de una activación neuronal, la extracción de oxígeno aumenta en la zona adyacente a las neuronas activadas provocando un cambio en la proporción entre oxiHb y deoxihemoglobina a favor de la deoxihemoglobina y, por lo tanto, provocando una caída inicial de la señal. Esta disminución de la señal inicial se ha observado en estudios de laboratorios que utilizan altos campos magnéticos (más de 3T). Milisegundos después se producen los cambios hemodinámicos descritos previamente, en que existe un incremento del flujo sanguíneo y del volumen sanguíneo local. Como resultado final existe un aumento progresivo del aporte de oxiHb que se produce tras 6-9 segundos del inicio de la activación. En este período la proporción oxi-deoxihemoglobina vuelve a cambiar, pero esta vez a favor de la oxiHb, lo que provocará un incremento de señal en el área determinada si se compara con la fase de reposo. Este fenómeno es el que se denomina efecto BOLD. La magnitud de esa señal aumenta en general de un 2 a un 3 % utilizando campos magnéticos de 1,5T (fig. 1.12).

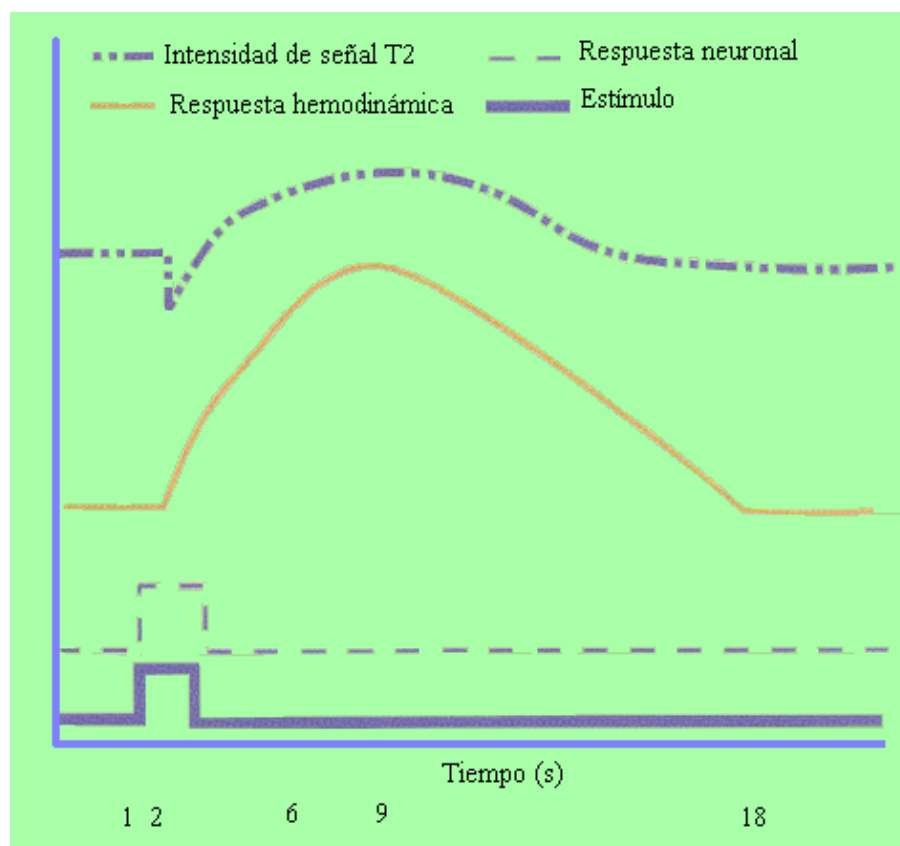


Figura 1.12: Evolución en el tiempo de los cambios neuronales, hemodinámicos y de señal de resonancia magnética.

FUNCIÓN CEREBRAL Y RESONANCIA MAGNÉTICA: RESONANCIA MAGNÉTICA FUNCIONAL

El efecto BOLD se utiliza para poder adquirir imágenes de fMR. En general se efectúa una adquisición de imágenes secuenciales mientras se realiza un ejercicio que suele constar de un tiempo durante el cual se ejecuta una tarea y un tiempo denominado de reposo. Posteriormente se comparan las imágenes obtenidas durante la activación cerebral con las de reposo y se calculan las diferencias en la señal de cada vóxel determinando si ha existido una variación en su intensidad durante la fase de activación en la región de interés generando un mapa de la actividad cerebral.

Así pues, la resonancia magnética tiene unas características que permiten estudiar la activación cerebral en tiempo real. La fMR tiene muchas más ventajas que otros métodos de mapeo de función cortical. Es una técnica no invasiva y por lo tanto no se asocia a los riesgos y complicaciones inherentes a un procedimiento agresivo. Si se compara con otros métodos no invasivos como el electroencefalograma, los potenciales evocados o la magnetoencefalografía, la fMR tiene una mayor resolución espacial y sobre todo tiene la capacidad de correlacionar el mapa funcional con la anatomía cerebral.

1.4. Diseño de tareas para Resonancia Magnética Funcional

Un aspecto fundamental en los estudios con fMR es el diseño de los experimentos. Es en este punto donde el diseño de los paradigmas empleados adquiere un papel central. Sin un diseño adecuado de la tarea no hay experimento, y sólo un profundo conocimiento de los procesos cognitivos implicados y sus posibles interacciones permitirá diseñar tareas de las que extraer resultados concluyentes [8].

El diseño de experimentos y su análisis debe permitir el conocimiento preciso de la actividad cerebral y sus partes constituyentes, e ir más allá de la mera localización de las áreas implicadas en una u otra actividad. A medida que van existiendo más hipótesis sobre la relación entre cognición y cerebro, irá aumentando el número de estudios guiados por hipótesis acerca del funcionamiento de los procesos, y por teorías cada vez más sólidas.

En líneas generales un diseño experimental debe permitir el control de, al menos, las siguientes variables:

1. Comparaciones experimentales.
2. Propiedades de los estímulos.
3. Tiempos de presentación.
4. Instrucciones para el sujeto.

La elección óptima de un paradigma experimental para resonancia magnética se basa en un equilibrio entre las limitaciones metodológicas y las propiedades de la fisiología que se pretende medir.

El objetivo del diseño será por tanto doble. Por un lado, determinar si un área está activa o no (capacidad de detección), y por otro, estimar el cambio de esa actividad a lo largo del tiempo (capacidad de estimación).

En este apartado se revisarán algunos de los paradigmas que están siendo empleados actualmente en el contexto de la fMR. De forma más específica, el diseño efectivo de paradigmas, las circunstancias bajo las que se debe escoger un tipo de diseño u otro, algunos aspectos a favor y en contra de cada uno de ellos así como algunas de sus dificultades.

CUESTIONES GENERALES: COMPARACIÓN DE DIFERENTES CONDICIONES EN RESONANCIA MAGNÉTICA FUNCIONAL

Los estudios en fMR se basan en la comparación de diferentes tareas o condiciones. Esta metodología tiene como característica la asunción de la denominada *inserción pura* en la que se basa la *sustracción cognitiva*.

Sustracción cognitiva y el problema de la inserción pura

La lógica de la *sustracción cognitiva* fue desarrollada por Donders en el siglo XIX, con el estudio de los tiempos de reacción para predecir procesos cognitivos subyacentes ([9]). De acuerdo con este procedimiento, se comparan dos eventos que difieren en una sola característica o factor. Por ejemplo, consideremos los tiempos de reacción de dos condiciones experimentales diferentes. En la condición A, el sujeto debe pulsar un botón siempre que se encienda una luz de color. En la condición B, el sujeto debe pulsar sólo si la luz es roja. Si asumimos que las tareas A y B difieren sólo en la discriminación del color de la luz, la resta de la tarea A menos la B revelará el tiempo que se tarda en hacer la discriminación del color. De este modo es posible obtener la duración de un evento difícilmente medible de forma directa. Este hecho se basa a su vez en la *inserción pura*.

La *inserción pura* implica que se puede sumar un proceso cognitivo a un conjunto de procesos ya activos sin afectarlos. Supongamos, por ejemplo, que durante la ejecución de una tarea X están activos una serie de procesos. A continuación deseamos hacer más compleja la tarea y añadimos algún componente nuevo a la misma. En este caso, la inserción pura asume que los procesos implicados en la tarea X no se han modificado, y sólo se han sumado unos nuevos mecanismos susceptibles de ser estudiados de forma independiente. Este hecho se da tanto en paradigmas experimentales de la psicología clásica, como en el campo de la neuroimagen.

Si la asunción de la inserción pura falla, la actividad obtenida podría no reflejar los cambios relacionados con el proceso de interés, sino con la interacción del nuevo componente cognitivo con los ya activados previamente. De no ser cierta la inserción pura existirían interacciones difíciles de estudiar, lo que lleva a la aparición de algunos problemas.

La *sustracción cognitiva* es la base de los procedimientos en fMR [10]. En neuroimagen, los niveles de actividad asociados a una tarea deben ser considerados en comparación con otra condición. Así, el nivel de señal está relativamente vacío de significado por sí solo. De esta forma, para que los datos en neuroimagen tengan sentido es necesaria la sustracción cognitiva. Las implicaciones no son banales, ya que lo que se postula es que la diferencia entre dos tareas puede ser considerada como un proceso cognitivo separado o un componente sensoriomotor diferenciado. Su aislamiento mediante este procedimiento permite la identificación del área especializada funcionalmente y los cambios en la respuesta hemodinámica.

Por tanto, una cuestión importante es la elección de las tareas a comparar. En este sentido se pueden caracterizar al menos dos aproximaciones diferentes:

1. Una concepción *dura* postula el control de tantas variables extrañas como sea posible, con el objetivo de aislar al máximo el componente cognitivo de interés. Así, las tareas a comparar serán lo más parecidas posible, difiriendo exclusivamente en el componente de interés.
2. Una concepción *blanda* señala que es preferible comparar la tarea de interés con una línea base de bajo nivel, lo que permitirá estudiar toda la red implicada. El caso más extremo de este tipo de diseños es comparar la actividad cerebral asociada a una tarea frente al reposo total (condición de control en la que se pide al sujeto que no haga nada).

Establecimiento de la condición de control

La elección de la condición de control es una de las tareas más importantes del investigador, ya que debe seleccionar aquella tarea que elimine del análisis todo lo que no sean componentes de interés.

El establecimiento de esta condición de control, incluso una línea base simple puede no ser tan sencilla. En numerosas ocasiones se establece como control un período de reposo, pidiendo al sujeto que no haga nada. Sin embargo, el sujeto puede estar realizando una tarea cognitiva y el cerebro sigue estando activo. El ejemplo más claro puede ser en estudios de memoria, en los que el sujeto debe aprender un material en los bloques de activación, pero no realizar ninguna tarea en los de reposo, sin embargo es muy difícil que el sujeto no repase las palabras. En este caso, se sugiere la necesidad de que el sujeto realice una tarea distractora que impida el repaso del material aprendido. Así, la tarea de control no desempeña sólo un papel como línea base con la que comparar, sino que en determinadas ocasiones funciona también como tarea distractora.

Es posible que durante la tarea de control (más aún si se trata de un período de reposo), cada sujeto tienda a mantener una línea de pensamiento sobre cualquier cosa de acuerdo con su estado mental subyacente. Por ejemplo, un paciente con trastorno obsesivo-compulsivo desarrollará unos pensamientos obsesivos, un paciente con depresión tendrá ideas depresivas, y sujetos normales pensarán en distintos aspectos de una forma no patológica.

Procesamiento cognitivo durante la condición de control y reposo

Ocurre en muchos experimentos que la activación cerebral es mayor en la condición de control que en las condiciones experimentales. Este resultado aparece en los resultados como una *desactivación*. Ésta es una cuestión objeto de investigación y difícil de

interpretar.

Algunos autores señalan que el cerebro presenta un procesamiento asociado a la línea base, espontáneo y virtualmente continuo, que funciona automáticamente y de forma sostenida, y sólo disminuye su nivel de activación con la presencia de alguna tarea relevante a la que se presta atención [11]. Su función podría estar relacionada con mecanismos automáticos muy relevantes para la supervivencia, como el rastreo y control del entorno, algunos mecanismos atencionales, la integración de procesos cognitivos y emocionales o incluso el procesamiento de pensamientos independientes de la estimulación externa. Sin embargo, no se dispone de un conocimiento preciso de los procesos implicados en el reposo, lo que es de gran importancia para que las interpretaciones de los resultados obtenidos sean correctas. Diversos estudios concluyen la existencia de una serie de áreas que muestran una actividad permanente durante diversos estados de reposo, y podrían incluir las cortezas mediales posteriores, la corteza lateral posterior, la corteza ventral prefrontal medial y la corteza dorsolateral prefrontal medial [12, 13].

TIPOS DE DISEÑO: CLASIFICACIÓN

Una distinción fundamental en el tipo de diseños es: diseños de bloques y diseños ligados a eventos (ER, *event-related*).

En el caso más sencillo de un diseño para fMR, los ensayos se agrupan en bloques de reposo y en bloques con la tarea de interés. El bloque diseñado para evocar un proceso cognitivo específico es el bloque *experimental* o de interés. Éste se alternará con los bloques denominados de control, que estarán diseñados para que evoquen todos los procesos cognitivos presentes en el bloque de actividad excepto el proceso de interés. De esta forma, las diferencias en la actividad neural entre las dos condiciones serán atribuibles al proceso objeto de estudio.

Los diseños experimentales para fMR se han ido haciendo más complejos, apareciendo diversas modalidades que son:

Diseños libres del contexto o de intersección

Estos diseños pueden tomarse como una extensión de la sustracción cognitiva. Sin embargo, tratan de controlar la posibilidad de interacciones entre distintos procesos y reducir el peso de la inserción pura. Así, se realiza el experimento en varias ocasiones y se contrasta el proceso cognitivo de interés con diferentes tareas control. Las áreas de activación comunes obtenidas de los diferentes contrastes se pueden atribuir al proceso que es objeto de estudio. De esta forma se emplea un método estadístico para determinar aquellas áreas que muestran actividad *libre del contexto*.

Estos diseños tienen el privilegio de no requerir que las tareas de control tengan en común con la tarea experimental *todo menos el proceso de interés*.

Diseños paramétricos

Tratan de identificar relaciones (generalmente lineales) entre la señal y el nivel de un parámetro de interés, del cual se presentan distintos niveles o grados. La premisa fundamental es que la fisiología regional cambiará sistemáticamente con el grado de implicación del proceso cognitivo de interés 1.13.

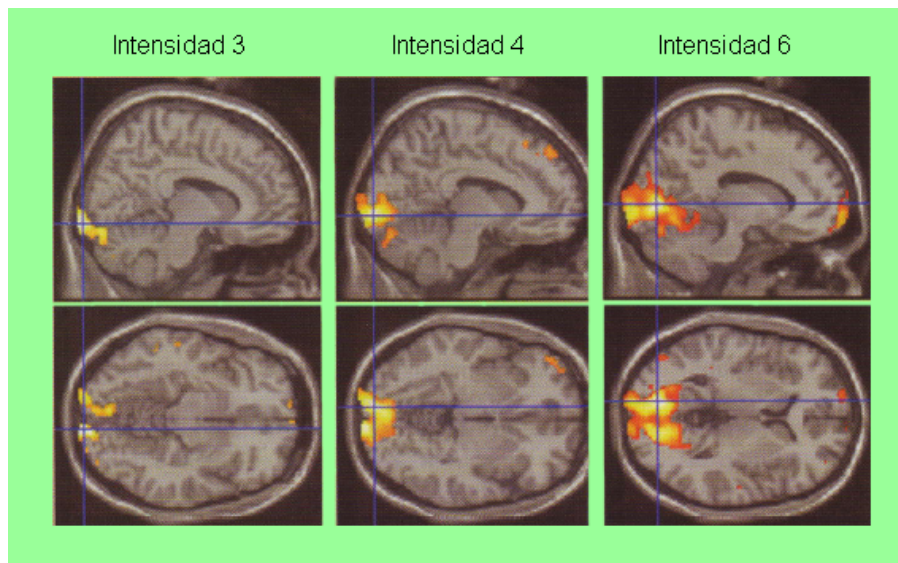


Figura 1.13: Diseños paramétricos. Influencia de los cambios en la intensidad de una fuente luminosa sobre la actividad en la corteza visual. A medida que aumenta la intensidad del estímulo (de intensidad 3 a 6), se producen cambios en el patrón de actividad en corteza visual [15].

Estos diseños son muy interesantes, por ejemplo, cuando se estudian los efectos de determinadas sustancias, con pruebas neuropsicológicas y rendimiento en tareas cognitivas [14].

Diseños factoriales

Estos diseños se están haciendo más frecuentes ya que permiten estudiar los efectos de un factor sobre otro, es decir, las interacciones entre distintos procesos [16]. Éstas, desde el punto de vista de los estudios cognitivos, se pueden interpretar como la integración de dos o más procesos cognitivos. Por ejemplo, la existencia de una interacción aparece si la diferencia de señal entre la presencia y ausencia del proceso A es diferente en sí misma cuando el proceso B está presente o ausente (de forma esquemática $[A + B + X] - [B + X]$ es diferente de $[A + X] - [X]$).

A continuación se describen diferentes modos de presentación de los estímulos. En

función de esta característica hablaremos de diseños de bloques, de diseños ER, o de diseños mixtos.

DISEÑOS DE BLOQUES

El diseño más utilizado en Resonancia Magnética Funcional para el estudio de la función cognitiva es el diseño de bloques. Se presenta una serie de ensayos con la misma tarea, agrupados en un período relativamente largo, denominado *bloque*. Estos períodos se alternan con bloques de *control*. Los bloques de control pueden ser: a) bloques de reposo, en los que el sujeto no realiza ninguna tarea y se registra la actividad basal, o b) bloques de activación, que implican la realización de otra tarea cognitiva 1.14.

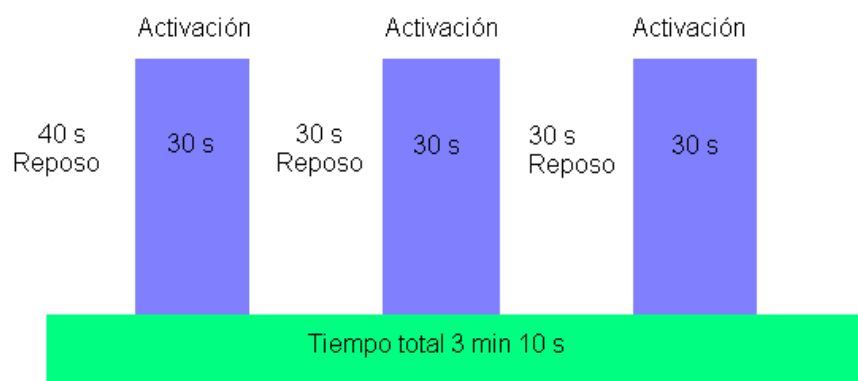


Figura 1.14: Diseño de bloques.

El diseño de bloques implica una resta o sustracción de la actividad encontrada en el período de activación y en el de reposo. El análisis básico muestra las áreas que presentan una diferencia de activación significativa estadísticamente entre las dos condiciones. Los bloques de cada secuencia se repiten de forma alternativa o aleatoria durante el experimento.

Los diseños de bloques son de gran interés cuando el objetivo del experimento es la detección de la activación, y permiten profundizar en la localización de actividad ligada a la diferencia entre las tareas presentadas.

Características o parámetros de un diseño de bloques

La *duración de los bloques* es uno de los principales parámetros a tener en cuenta. En general, los bloques más largos permiten la estabilidad de respuestas sostenidas en el tiempo, lo que facilita que la respuesta hemodinámica se sature alcanzando un máximo, hecho que ocurre aproximadamente tras 10 segundos de estimulación. Por otro lado, los bloques más cortos permiten incrementar el número de transiciones entre las condiciones

sin aumentar la duración del estudio, de forma que aumenta la variabilidad asociada a la tarea (relativa a la condición de control). También, es importante considerar la *periodicidad de los bloques*, ya que puede incrementar la varianza de los datos.

Otro parámetro de interés es la *duración de los estímulos*. En los diseños de bloques, para aumentar la sensibilidad del estudio, hay que maximizar el tiempo en el que el estímulo está presente y minimizar el intervalo entre estímulos (ISI, *interstimulus interval*).

El *número de bloques* es otra variable de interés al diseñar el experimento. Cuanto mayor sea el número de bloques, mayor será la potencia de contraste. Sin embargo, una gran cantidad de bloques aumentará la duración del estudio y por extensión, todos los efectos negativos asociados al cansancio del sujeto, a la habituación de la tarea, etc. En este caso, se mejora la calidad estadística del estudio a costa de su validez psicológica. Por el contrario, un diseño con pocos bloques es más sensible a la deriva o inestabilidad de la señal de la resonancia magnética.

Ventajas e inconvenientes de los diseños de bloques

Los diseños de bloques presentan algunas ventajas que los hacen especialmente útiles en determinadas circunstancias. Dichas ventajas son:

- Buena relación señal-ruido.
- Rápidos de realizar/Registros rápidos.
- Elimina el cambio de tarea rápido y constante propio de los diseños event-related.
- Particularmente útiles en: componentes cognitivos de interés que no están sujetos a variación paramétrica; todos los ensayos son categóricos y consistentes; estimulación pasiva o sensorial posible.
- Algunas manipulaciones experimentales pueden tener más efecto en diseños de bloques.
- Algo que caracteriza a los diseños de bloques, y es de gran valor, es la mayor potencia estadística/potencia de contraste que ofrecen.

Sin embargo, este tipo de diseños adolece de algunas limitaciones:

- Muy sensibles a la deriva de señal: sensibles al movimiento de la cabeza, especialmente cuando se usan pocos bloques.
- Una mala elección de la línea base puede llevar a conclusiones incorrectas.
- No se pueden realizar muchas tareas de forma simultánea o repetida.
- Dificultad para identificar la magnitud de la activación.

- Incrementa la fatiga y los cambios en el nivel de atención.
- Facilita la anticipación de los sujetos (predictibilidad de los acontecimientos).
- Dificultad para mantener el mismo set cognitivo durante los bloques largos, debido al aburrimiento y la pérdida de interés en la tarea.
- Habituaación rápida con determinados estímulos.
- Asume la inserción pura y la sustracción cognitiva.
- Asume la consistencia entre los ensayos (las respuestas cognitivas pueden diferir en tiempo de repetición y otras características).
- Imposibilidad de separar y estudiar respuestas individuales.
- Dificultad de encajar en diseños de bloques determinados paradigmas (p.ej., paradigmas oddball, donde el interés reside en el efecto novedoso del estímulo al romper el contexto anterior).

DISEÑOS LIGADOS A EVENTOS

Los diseños ER permiten aplicar a la resonancia magnética una metodología que ha sido empleada con éxito en otros campos de la psicología y la neurociencia. Se trata de aplicar la técnica de detección de respuestas a ensayos individuales, de una forma más o menos análoga a los estudios con potenciales evocados relacionados con eventos registrados mediante electroencefalograma [17]. Estos procedimientos permiten la adquisición de las imágenes durante los cambios transitorios asociados a eventos cognitivos, sensoriales o motores individuales, posibilitando la caracterización de subetapas implicadas en los diferentes procesos.

Los diseños de fMR-ER se diferencian de los de bloques en que se registran las respuestas ante eventos discretos y únicos en lugar de registrar la integración de señales durante una cierta cantidad de tiempo, correspondiente a muchos ensayos. Esto permite la presentación de las condiciones de forma aleatoria, lo que previene que la media de las respuestas no se vea modulada por un contexto específico o por una tendencia de respuestas. De este modo se evitan algunos de los problemas de los diseños de bloques, como los derivados de las expectativas y la habituación.

Características y parámetros de un diseño ligado a eventos (ER)

Los diseños ER no son un diseño único. La modalidad de presentación de los estímulos hace que puedan ser divididos, al menos, en dos tipos: diseños ER no solapados (*periodic single trial*) y diseños ER inquietos (*jittered single trial*) (fig. 1.15). La denominación “inquieto” hace referencia a la variabilidad en cuanto a los momentos de aparición de los ensayos (se mantiene constante el tiempo medio de aparición, pero la aparición de

cada ensayo individual es impredecible), y en cuanto al tipo de ensayo, ya que la aparición de las diferentes condiciones del estudio se aleatoriza.

El parámetro que condiciona estos tipos de diseños es la elección del ISI, o de forma más precisa el tiempo transcurrido desde el inicio de un estímulo hasta el inicio del siguiente estímulo (SOA, *stimulus onset asynchrony*). La elección del intervalo depende de las características del experimento, pero como norma general es preferible un SOA corto y distribuido aleatoriamente a lo largo del experimento. Las tasas rápidas de presentación de los ensayos permiten el mantenimiento del set atencional, y disminuye la posibilidad de distracciones del sujeto.

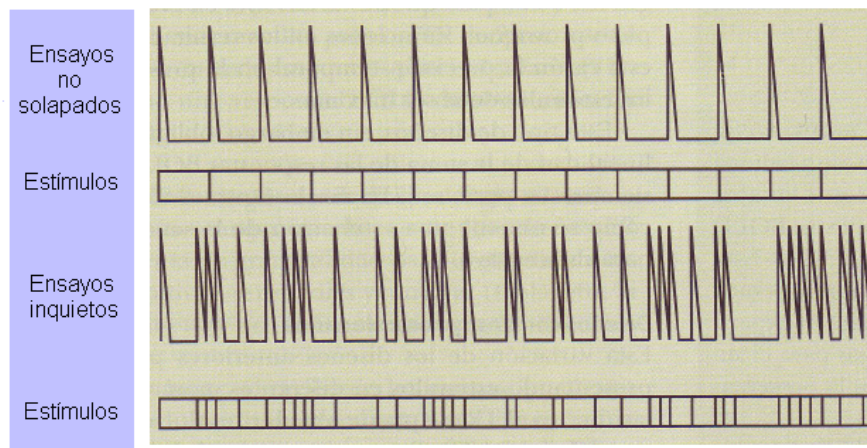


Figura 1.15: Tipos de diseño ligados a eventos.

Diseños ligados a eventos no solapados: Es el caso más sencillo de diseño ER (ver fig. 1.15). Los ensayos se presentan de forma poco frecuente con SOA muy largos. Éste debe permitir que la HDR (high dynamic range) descienda hasta la línea base y no se solape con la HDR del siguiente ensayo. Para ello, el espacio entre ensayos parece ser óptimo (máxima sensibilidad) hacia los 16 s.

Diseños ligados a eventos inquietos: Estos diseños surgen al observar que se pueden presentar numerosos ensayos en una rápida sucesión, y estudiar las respuestas BOLD superpuestas relacionadas con cada ensayo, sin necesidad de esperar a que la HDR regrese a la línea base. Para ello, se varía la cadencia en la presentación de estímulos durante la adquisición. Esto permite realizar experimentos de forma mucho más rápida.

Ventajas e inconvenientes de los diseños ligados a eventos

Existen algunas ventajas que hacen de los diseños ER una herramienta de gran utilidad para el estudio de los procesos cognitivos, sensoriales y motores:

- Permiten estudiar determinados procesos difíciles de encajar en un diseño de bloques.
- Flexibilidad y presentación aleatoria de los ensayos: elimina la predictibilidad de los diseños de bloques.
- Permiten la agrupación a posteriori de los ensayos de acuerdo con determinadas variables como la respuesta de los sujetos (correctas-incorrectas; recordadas-olvidadas, etc.)
- Se pueden estudiar los acontecimientos raros e impredecibles. Incluso permite el estudio de acontecimientos que sólo pueden ser indicados por el propio sujeto experimental.
- Permiten estudiar la dinámica temporal de la respuesta.

Pese a sus grandes ventajas, los diseños ER presentan también algunas dificultades:

- Algunos procesos no pueden ser estudiados en períodos cortos de tiempo, y es necesario estudiarlos de forma sostenida (por ejemplo, la atención sostenida).
- En general, estos diseños son menos eficientes para detectar un efecto que los diseños de bloques.
- Algunas de sus ventajas no son siempre aplicables, como la presentación aleatoria de estímulos (por ejemplo, el estudio de la memoria operativa).
- Se produce un efecto diferencial de los intervalos interestímulos. Así, los intervalos largos no incrementan de forma óptima la varianza del estímulo, y los intervalos cortos pueden producir efectos refractarios.
- Puede haber dificultades con determinadas poblaciones. Por ejemplo, la presentación aleatoria de las tareas obliga al cambio de tarea constantemente. En este caso, niños pequeños o pacientes con lesiones frontales pueden no realizar correctamente las tareas, precisamente por el cambio.

DISEÑOS MIXTOS

Son el resultado de una combinación de los diseños de bloques y de eventos. Estos diseños incorporan bloques de ensayos variables que permiten distinguir entre el procesamiento sostenido y el procesamiento instantáneo. La hipótesis con la que se trabaja es que determina dos elementos de la tarea son sostenidos en el tiempo, mientras otros son específicos para cada estímulo y se modulan en función de éste. Es decir, algunos procesos son elicitados por el objetivo global de la tarea (*estado*), mientras otros procesos aparecen y desaparecen con los estímulos presentados (*respuestas al ítem*).

Así, este tipo de diseños combina dos características, la capacidad para detectar efectos dependientes del estado, clásica de un diseño de bloques y, la de los ER, en la que se evalúan los efectos relacionados al ítem. Los análisis se realizan de forma independiente para ambos tipos de medida.

Un buen ejemplo de diseño mixto es el de Otten et al. [20]. Estos autores muestran cómo el éxito en una tarea de recuerdo viene modulado no sólo por la actividad fásica, sino también por el tono general durante la tarea (actividad sostenida). En este estudio, los autores explican cómo los sujetos entraban en una especie de *modo* o *estado de recuerdo* que explica parte de los resultados conductuales obtenidos en la tarea de recuerdo. Otro ejemplo muestra la interacción entre atención selectiva y los cambios transitorios en el estímulo [21]. Los sujetos veían un estímulo que ocasionalmente cambiaba de color o el movimiento. En algunos bloques los sujetos debían detectar los cambios de color, mientras que en otros debían responder a los cambios en el movimiento. Los análisis en relación con los bloques mostraron que la atención a un atributo visual específico del estímulo (por ejemplo, el color) incrementaba la línea base en las regiones específicas para procesar ese atributo. Por otra parte, el análisis de los ensayos individuales mostró que las respuestas de los sujetos aumentaban su amplitud cuando atendían a ese atributo específico.

1.5. Preprocesado en Resonancia Magnética Funcional

Se denomina preproceso al tratamiento digital que se aplica sobre las imágenes de fMR, una vez generadas por el escáner, y que persigue la mejora de la calidad de las imágenes. De esta forma se minimizan las fuentes de error que pudieran sesgar el análisis estadístico que se realice a continuación.

El preproceso puede tener diferentes etapas, con mayor o menor sofisticación matemática, como por ejemplo: mejorar la relación señal/ruido, corregir determinados artefactos intrínsecos y normalizar las imágenes. La normalización, en particular, permite establecer comparaciones entre sujetos. Una vez mejorada la calidad de las imágenes, el análisis estadístico, requiere varios niveles de estudio: nivel de sujeto, que a su vez se subdivide en dos como son la detección de actividad para cada tarea y la comparación entre tareas; nivel de grupo, lo cual requiere la normalización intersujetos y la comparación entre tareas de distintos sujetos, y por último, un tercer nivel de análisis, donde se comparan diferentes grupos (normales frente a pacientes, hombres frente a mujeres, etc).

A continuación vamos a describir, brevemente, los pasos más habituales de los que consta el preproceso.

REGISTRO

Para hacer un seguimiento de la fisiología es necesario realizar numerosas adquisiciones a lo largo del tiempo. Debido a la rapidez de la adquisición, con frecuencia, las imágenes se desplazan unas con respecto a otras. El registro consiste en un realineado geométrico de todas las imágenes mediante transformaciones rígidas afines con 6 grados de libertad (3 traslaciones y 3 rotaciones). En caso de no hacerse aparecen falsos positivos y negativos fundamentalmente en bordes de la imagen como en cráneo, límite de los ventrículos, etc.

UNWRAPPING

En las imágenes funcionales adquiridas mediante la secuencia EPI, debido a los rápidos cambios de fase que se producen en la secuencia y a la baja intensidad de señal se producen pérdidas significativas en la imagen, en particular en zonas óseas, de carácter no lineal, y, además, con una apreciable merma en el tamaño de la imagen funcional respecto de la imagen anatómica. Este fenómeno es también una fuente potencial de falsos positivos que ha de ser corregido mediante las técnicas de *unwrapping*. En la figura 1.16 podemos ver, en corte axial, la diferente calidad de imagen entre aquellas adquiridas mediante una secuencia tipo espín eco (imagen anatómica), y las mismas imágenes adquiridas mediante EPI de eco de gradiente. La máxima señal de BOLD puede observarse en las imágenes adquiridas mediante la secuencia EPI, sin embargo, aparecen deformaciones y pérdidas de señal llamativas en detalles más estructurales o morfológicos.

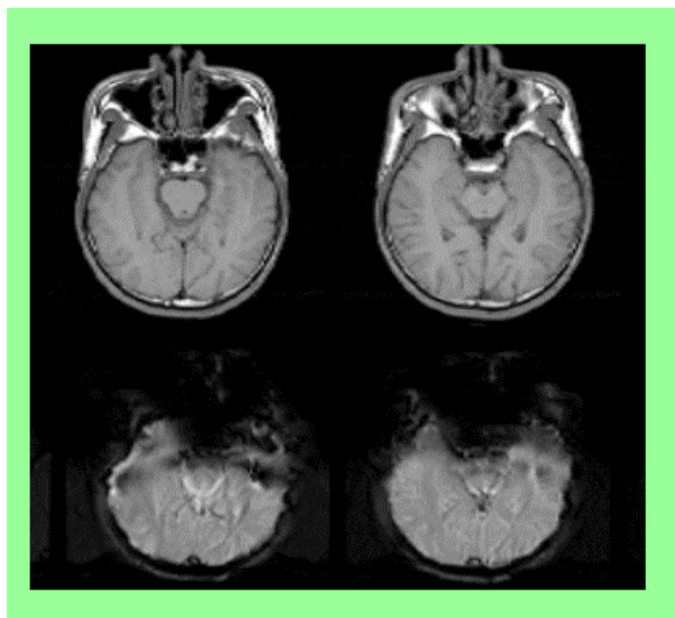


Figura 1.16: Comparativa de calidad de imagen entre espín-eco y eco de gradiente para los mismos cortes.

En la figura 1.17 podemos observar, también sobre un corte axial, las inhomogeneidades detectadas en imágenes obtenidas con EPI. La figura 1.17A contiene una imagen obtenida con una antena diseñada para una mejor detección de actividad en el córtex visual. Esta imagen contiene una doble inhomogeneidad. Por un lado debida al tipo de antena, y otra debida a la propia secuencia EPI de eco de gradiente. La figura 1.17B muestra un estudio de inhomogeneidad (realizado con SPM) para el corte presente en la figura 1.17A. En él se analizan las inhomogeneidades a lo largo de diferentes direcciones en el interior de la imagen. Con estos estudios de inhomogeneidad se pueden construir filtros que permitan paliar este defecto para cada uno de los cortes.

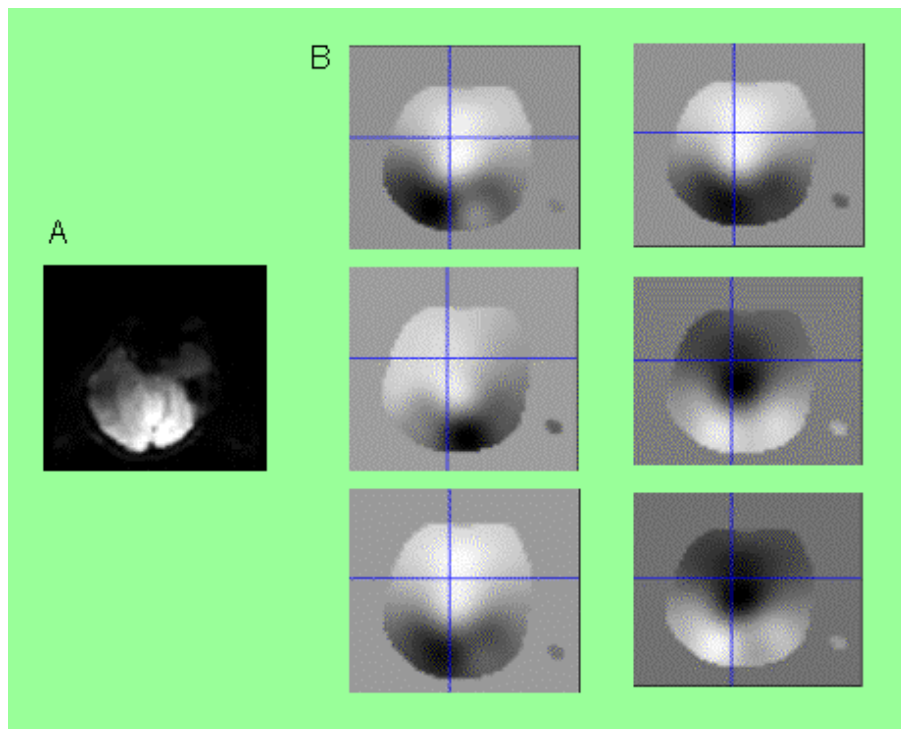


Figura 1.17: Estudio de la inhomogeneidad y distorsión en la secuencia EPI mediante el paquete SPM. A) Imagen EPI eco de gradiente real. B) Distorsión a lo largo de diferentes direcciones.

CORREGISTRO

En la figura 1.16 se ha observado la diferente información y calidad entre una imagen anatómica y una imagen funcional. Es muy útil poder hacer coincidir ambas informaciones en una única representación visual. Por tanto, se denomina corregristo, al registro de imágenes provenientes de diferentes modalidades de imagen, y/o de la misma modalidad pero con información anatómica y funcional respectivamente.

CORRECCIÓN DE TIEMPO DE CORTE

Un elemento clave en el análisis estadístico posterior es la correspondencia entre las épocas de la tarea (activación y reposo) y la secuencia de volúmenes adquiridos. Para la realización del análisis estadístico posterior hay que agrupar dichos volúmenes por cada época. En la práctica, cada corte de un volumen es adquirido en instantes de tiempo diferentes aunque sucesivos, por lo que no se corresponde con el mismo instante en la ejecución de la tarea por parte del sujeto. Por medio de interpolaciones matemáticas, se puede regenerar el estudio de forma que los cortes de un mismo volumen sean muy similares a una adquisición ideal en la que todo un volumen se adquiriese en el mismo instante de tiempo.

FILTRADO PASO BAJO

Las tareas suelen ser repetitivas con objeto de adquirir varias muestras en el mismo estado del sujeto (ya sea en reposo o activación). Esto significa que la actividad ha de aparecer y desaparecer con una frecuencia cercana a la frecuencia de repetición de los ciclos de activación-reposo. Por este motivo se filtran en frecuencia componentes en la señal que tengan frecuencias superiores a la frecuencia de repetición de la tarea. De esta forma, se consigue eliminar ruido de alta frecuencia y algunos artefactos.

ELIMINACIÓN DE TENDENCIAS LINEALES (LINEAR TREND REMOVAL)

El estado basal puede aumentar o disminuir a lo largo de la tarea. Esto hace que podamos tener diferencias significativas entre la amplitud de la señal al principio de la tarea con respecto al final de ésta. Estas tendencias aumentan las proporciones de falsos positivos y negativos dependiendo también de la forma en que se realice la estadística. Conviene tratar de detectar estas tendencias para su eliminación previamente a cualquier tipo de análisis estadístico.

SUAVIZADO

Es habitual la utilización de un filtrado gaussiano denominado suavizado. Hay numerosas razones que aconsejan hacerlo. En primer lugar para mejorar la relación señal/ruido. Es decir eliminar ruido, tanto en la imagen como en la señal a lo largo del tiempo. El suavizado tiende a difuminar la imagen: detalles, bordes, etc. Pero esto, lejos de ser un problema es, sin embargo, una virtud en imagen funcional, debido a que los tamaños de la actividad suelen ser mayores que la extensión del suavizado. En la práctica, en muchos casos, se opta por no aplicar suavizados y, sin embargo, realizar adquisiciones cuyo tamaño de píxel sea mayor para que la relación señal ruido mejore de forma intrínseca a la adquisición.

NORMALIZACIÓN

Este es el punto más delicado del preproceso. Sólo es necesario cuando se desea realizar comparaciones con otros sujetos o cuando se quiere asociar un área de la anatomía del sujeto con un área del atlas de Talairach. En la actualidad se emplean dos posibles métodos de normalización. O bien, el ajuste a la rejilla de Talairach, o bien, el ajuste al cerebro normalizado del Instituto Neurológico de Montreal (MNI, Montreal Neurology Institute). Talairach, al ser más antiguo, es el más referenciado en la literatura médica, sin embargo, dicho atlas se realizó sobre un cerebro de un determinado sexo por lo que dista bastante de lo que se entiende por un cerebro estándar o promedio. El cerebro MNI tiene la ventaja de ser el promedio de numerosos casos adquiridos tanto con resonancia EPI como con PET. En la figura 1.18 se muestra un ejemplo de ajuste de rejilla de Talairach a una adquisición anatómica.

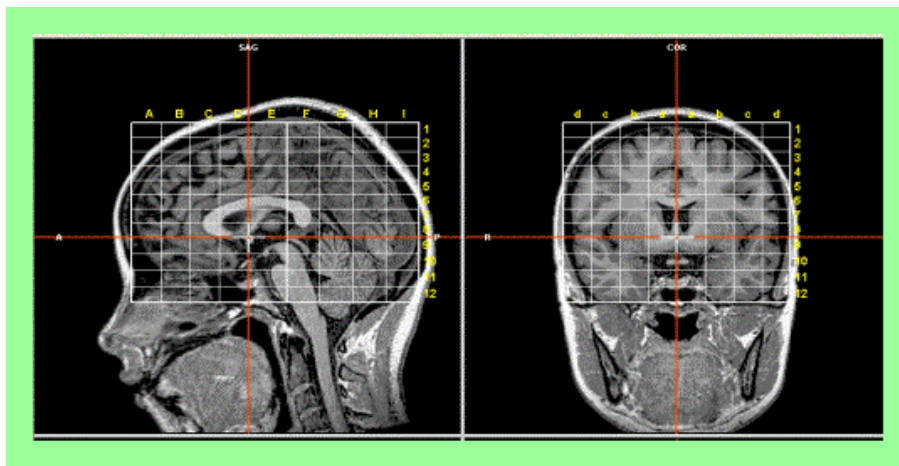


Figura 1.18: Ajuste de la rejilla de Talairach a imagen anatómica.

En la figura 1.19 vemos el cerebro normalizado MNI realizado como promedio de adquisiciones EPI (fig. 1.19A) y, a su derecha, vemos una adquisición anatómica de un sujeto ajustada geométricamente al cerebro MNI-EPI (fig. 1.19B). Este último ajuste es de gran importancia para poder corregistrar anatómico con funcional, dado que la imagen EPI debido a la escasa intensidad de la señal, tiene unas dimensiones ligeramente mermadas con respecto a una imagen obtenida mediante una secuencia anatómica.

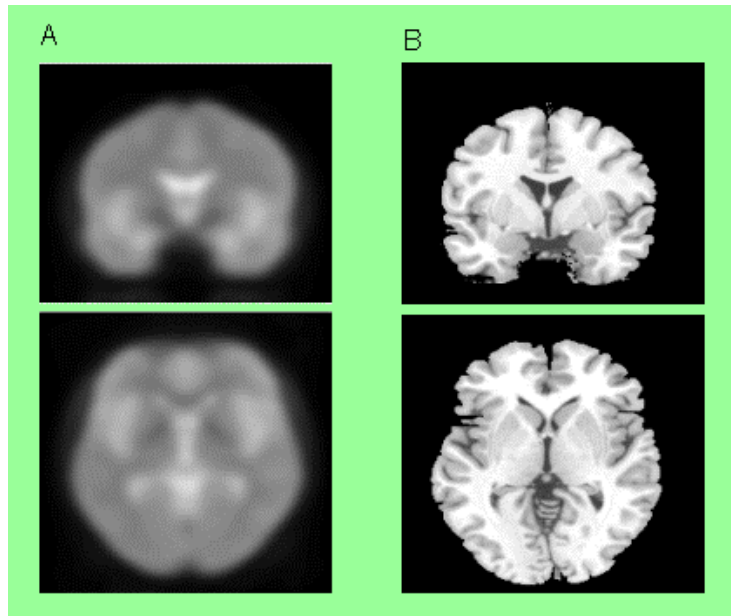


Figura 1.19: *Ajuste anatómico a cerebro normalizado MNI-EPI. A) Corte coronal y axial del cerebro MNI-EPI. B) Cortes coronal y axial de adquisición anatómica ajustada al cerebro normalizado MNI-EPI.*

Si normalizamos, tanto las imágenes anatómicas como las funcionales, a MNI o a Talairach, posteriormente podemos corregistrar ambos tipos de imágenes con mucha mayor precisión. Es importante tener en cuenta que es fácil pasar de las coordenadas de Talairach a las coordenadas MNI mediante una sencilla transformación matemática. La normalización, de por sí compleja, puede serlo aún más si los sujetos del estudio tienen malformaciones o han sido intervenidos quirúrgicamente dando lugar a pérdidas de masa encefálica. En estos casos la normalización es un proceso casi manual y, por tanto, muy tedioso.

ANÁLISIS ESTADÍSTICO

Después del posproceso el problema estadístico consiste en localizar, en presencia de ruido, las regiones donde hay activación.

Como ya se ha indicado anteriormente, para cada sujeto del estudio se adquieren un gran número de imágenes básicamente en dos estados: una condición de control y una condición experimental con el proceso de interés. A partir de ahí es necesario saber en qué zonas del cerebro aparece una diferencia significativa entre la condición de activación y la condición de control. Estadísticamente requiere un test que puede ser implementado de diversos modos: correlación, t-student, ANOVA, modelo general lineal y, también, estadística paramétrica.

En la figura 1.20, obtenida con el paquete BrainVoyager, se muestra, para una zona de actividad, su serie temporal. De ésta se extrae la señal en las épocas de activación, que son promediadas, y también se extrae y promedia la señal en las épocas de reposo. La figura 1.20A muestra lo que sería el resultado final. La figura 1.20B muestra la serie temporal promedio de un conjunto de píxeles que provienen de una zona activada. La figura 1.20C contiene el promedio de la señal para las épocas de activación y el promedio para las épocas de reposo. La comparación de esas dos señales promedio, nos proporciona la información de si la señal es significativamente diferente entre ambos estados.

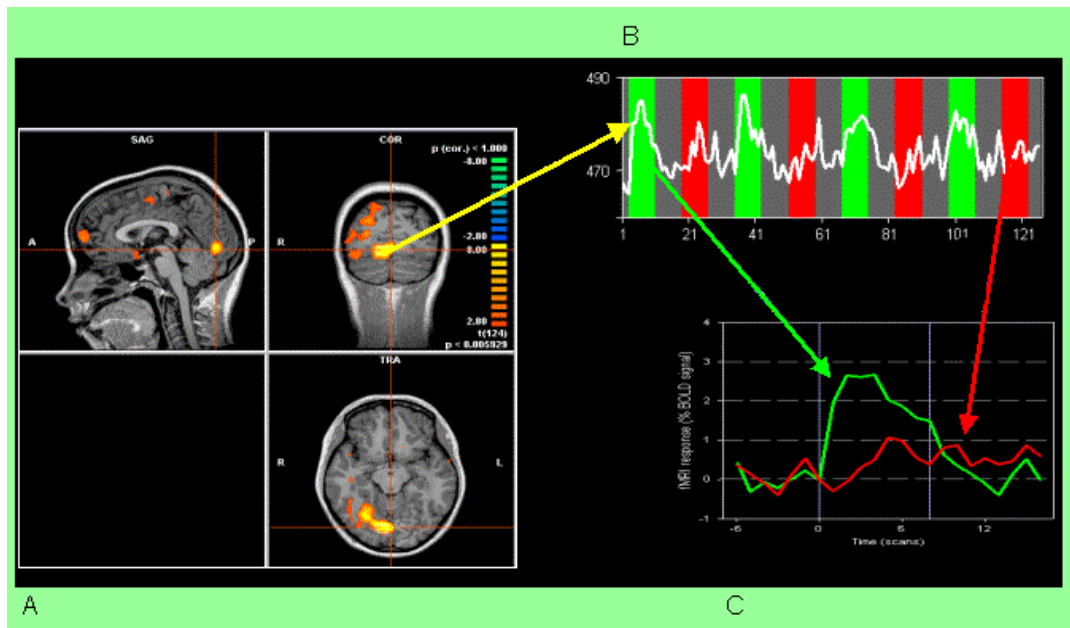


Figura 1.20: Promedio de épocas de activación y épocas de reposo para su posterior test estadístico. Comparación de actividad y reposo (A). Resultado final del estadístico para una $p < 0,006$ (B). Serie temporal promedio de un conjunto de píxeles activados (C).

1.6. Métodos univariantes: GLM

1.6.1. Modelo de datos

En un experimento fMRI se obtienen N bloques BOLD, esto es, una serie temporal de N imágenes tridimensionales que indican la variación de la oxigenación en el cerebro. Estas imágenes se almacenan a efectos de su representación visual en matrices tridimensionales con una ordenación siguiendo los tres ejes del espacio de la imagen. Sin embargo, a efectos de procesamiento de la señal, los valores de estas matrices (voxels) se redistribuyen en forma arbitraria en vectores $x[n]$ de dimensión M , donde $0 \leq n \leq N - 1$ es el índice temporal de la serie. La concatenación en columna de todos estos vectores, cada uno de los cuales se representa en formato fila, constituye la matriz de respuesta

hemodinámica \mathbf{X} , de dimensiones $N \times M$.

El experimento fMRI consiste usualmente en la presentación simultánea de estímulos sensoriales al sujeto en bloques e intervalos predefinidos, o bien en la realización de una actividad (motora, cognitiva...) a intervalos preestablecidos que se señalan mediante algún medio sensorial al sujeto. Los instantes de inicio y final de los estímulos y actividades en cada bloque se representan mediante unas señales indicadoras binarias, cada una de las cuales denominada vector de referencia o variable explicatoria. Cada una de estas L señales se filtra mediante un filtro gaussiano que aproxima la respuesta hemodinámica del cerebro a la actividad o estímulo en cuestión. El conjunto de los vectores de referencia filtrados $y_l[n], 0 \leq l \leq L - 1$ constituye la denominada matriz de diseño. En esta matriz se pueden incluir eventualmente actividades no relacionadas con el experimento, tales como estímulos adicionales, respiración, ritmo cardíaco, etcétera. Esta matriz se denotará como matriz de diseño \mathbf{Y} , de dimensiones $N \times L$.

1.6.2. Mapa paramétrico estadístico (SPM)

El método SPM supone la construcción de procesos estadísticos para probar hipótesis sobre efectos regionalmente específicos. Los mapas paramétricos estadísticos son imágenes cuyos voxels representan valores que están distribuidos de acuerdo con una función de densidad de probabilidad conocida, que puede ser una t de Student o una distribución F (mapas t o F), distribuidos en torno a la hipótesis nula. De esta forma, cuanto mayor sea un valor dado, más improbable es la hipótesis nula en ese voxel. Para estimar estos valores, se hace un análisis de todos los voxels usando un test estadístico y los parámetros estadísticos resultantes son reunidos en una imagen, llamada t -maps o F -maps.

Para analizar y hacer inferencias clásicas, este método utiliza conjuntamente el Modelo Lineal General (GLM) (Friston et al., 2006) y la Teoría de los Campos Aleatorios (Harville, 1977).

1.6.2.1. Modelo Lineal General

Todos los estadísticos paramétricos están basados en el GLM. Este modelo se diseña de forma que, para cada voxel, se estima linealmente la respuesta hemodinámica a partir de la matriz de diseño. Concretamente, para cada voxel m el modelo se puede expresar como:

$$x_m[n] = \mathbf{y}^T[n] \boldsymbol{\beta}_{m1} + \beta_{m0} + e_m[n] \quad (1.2)$$

donde $e[n]$ es el error de predicción, $\boldsymbol{\beta}_{m1}$ es un vector de L componentes correspondientes a otros tantos estímulos y/o actividades y β_{m0} es un factor de sesgo. Usualmente se incluye una componente constante en la matriz de diseño para absorber este sesgo. Alternativamente, se puede eliminar el sesgo de la serie temporal. Este método es univariable, es decir, la estimación de cada vector $\boldsymbol{\beta}_{m1}$ sólo depende de la respuesta hemodinámica en

el voxel m . Sin embargo, el problema puede resolverse en bloque, para lo que puede ser escrito en notación matricial de la siguiente forma:

$$\mathbf{X} = \mathbf{Y}\boldsymbol{\beta} + \mathbf{E} \quad (1.3)$$

donde \mathbf{X} es la respuesta cerebral e \mathbf{Y} es la matriz de diseño anteriormente definida. El conjunto de parámetros $\boldsymbol{\beta}$ es una matriz de dimensiones $L+1 \times M$. Las filas de esta matriz pueden ser reordenadas según el orden original de las matrices bloque conteniendo los bloques BOLD a efectos de su representación gráfica en el espacio. Las L primeras filas así reordenadas representan una imagen del cerebro cuyos voxels indican el nivel de presencia de cada uno de los estímulos en la correspondiente zona del cerebro. La última fila representa el sesgo de cada uno de los voxels en el estimador y , en principio, carece de interés neuroanatómico o clínico.

La matriz \mathbf{E} contiene el error de estimación por cada voxel e instante de tiempo. Se supondrá que los errores son independientes e idénticamente distribuidos con covarianza $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$.

Si la inversa de $\mathbf{Y}^T \mathbf{Y}$ existe, lo que ocurre cuando la matriz de diseño es de rango completo, las estimaciones de mínimos cuadrados son:

$$\hat{\boldsymbol{\beta}}_{GLM} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X} \quad (1.4)$$

1.6.2.2. Inferencia estadística

La hipótesis nula de que los efectos contenidos en \mathbf{Y} no son significativos se puede probar con el estadístico t usando componentes lineales o contrastes de las estimaciones de los parámetros $\hat{\boldsymbol{\beta}}$. Un contraste \mathbf{c} es un vector fila de pesos. La importancia de una composición lineal de efectos particulares es probada con:

$$T = \frac{\mathbf{c}\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}\epsilon^2(\mathbf{Y}^T \mathbf{Y})^{-1}\mathbf{c}^T}} \quad (1.5)$$

donde ϵ^2 es la estimación de la varianza σ^2 , la cual es estimada a cada voxel usando un estimador usual de mínimos cuadrados.

El uso del estadístico t no aporta ningún tipo de información aparte de la contenida en el parámetro $\boldsymbol{\beta}$ y no resulta de interés para la tarea de clasificación, aunque sí su visualización.

ALGORITMOS MULTIVARIANTES: SVM Y REDES DE PROCESOS GAUSSIANOS

2.1. Máquinas de Vectores Soporte (SVM)

El método de los vectores soporte (SV) es un método general para la resolución de problemas de clasificación, regresión y estimación. El método fue propuesto por V. Vapnik en 1964 en problemas de reconocimiento de patrones. En los años 90 el método fue generalizado y en la actualidad es objeto de un gran interés.

Las Máquinas de Vectores Soporte son nuevas estructuras de aprendizaje basadas en la teoría estadística del aprendizaje. Se basan en transformar el espacio de entrada en otro de dimensión superior (infinita) en el que el problema pueda ser resuelto mediante un hiperplano óptimo (de máximo margen) y de forma lineal:

$$y = (w \cdot x) + b$$

2.1.1. Introducción

Dada una tarea con un conjunto de datos de entrenamiento finito, la generalización del buen funcionamiento de la máquina cuando se introducen nuevos datos se lleva a cabo si hay un compromiso entre: la precisión o la mayor probabilidad de acierto dentro del conjunto de datos de entrenamiento y la capacidad (o complejidad) de la máquina, es decir la habilidad de la máquina a aprender cualquier conjunto de entrenamiento sin error [18].

Una máquina sobreentrenada: No cometerá errores en el conjunto de datos de entrenamiento, pero a la hora de realizar la fase de test (datos nuevos) cometerá muchos errores. Todo lo que no se ajuste rigurosamente a lo que ha aprendido en el entrenamiento lo clasificará mal.

Una máquina subentrenada: Está tan poco entrenada que no tiene información suficiente para poder clasificar bien en la fase de test.

En ninguno de estos dos casos la generalización de la máquina será buena.

Para un conocimiento más profundo que el que se aporta en este capítulo, consultar [19].

2.1.2. Límites de la generalización de una máquina de aprendizaje

Existen una familia de límites que gobiernan la relación entre la capacidad de la máquina y su desarrollo.

Suponiendo que se tienen l observaciones, cada observación consiste en un par: $x_i \in \mathbb{R}^n$, $i=1, \dots, l$ y su etiqueta asociada y_i , que define la clase a la que pertenece el dato de entrada. Por ejemplo, si el dato de entrada es el patrón que estamos buscando ($y=1$), si no ($y=-1$). La tarea de la máquina es aprender la correspondencia de y_i dado un dato de entrada x_i . Realmente, la máquina está definida por un conjunto de posibles correspondencias $x \mapsto f(x, \alpha)$, dada una entrada (x) y la elección de un parámetro (α) la salida será siempre la misma ($f(x, \alpha)$), por eso se dice que la máquina es determinista. Una elección particular de α genera lo que se llama una máquina entrenada.

El error en una máquina entrenada es:

$$R(\alpha) = \int \frac{1}{2} \|y - f(x, \alpha)\| dP(x, y) \quad \text{donde : } dP(x, y) = p(x, y) \cdot dx \cdot dy \quad (2.1)$$

A ese error se le llama riesgo esperado, riesgo real o simplemente riesgo; es desconocido porque no se sabe lo que vale $p(x, y)$.

El riesgo empírico: $R_{emp}(\alpha)$, es la tasa de error medido en el conjunto de entrenamiento (para un número finito de observaciones):

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l \|y_i - f(x_i, \alpha)\| \quad (2.2)$$

La complejidad de una máquina se puede medir en términos de la dimensión VC. La dimensión VC(h) para un conjunto de funciones $f(\alpha)$ es definida como el número máximo de puntos de entrenamiento que pueden ser clasificados por $f(\alpha)$. Si la dimensión es h , existen al menos un conjunto de h puntos que pueden ser clasificados de forma arbitraria, pero en general no todos los conjuntos de h puntos pueden ser clasificados.

Existe otro tipo de riesgo que depende de la capacidad (o complejidad) de la máquina, a este término se le llama confianza de Vapnik Chervonenkis (VC) [22], y depende de la dimensión VC. La confianza de Vapnik Chervonenkis se expresa como:

$$R = \sqrt{\frac{h(\log(\frac{2l}{h}) + 1) - \log(\frac{\eta}{4})}{l}} \quad (2.3)$$

donde: η toma valores comprendidos entre 0 y 1; h es la dimensión VC.

Con probabilidad $1 - \eta$ se puede mantener el siguiente límite del riesgo [23]:

$$R(\alpha) \leq R_{emp}(\alpha) + R \quad (2.4)$$

Sobre este límite se pueden señalar tres características:

- Es independiente de $P(x,y)$.
- No es posible hallar el lado izquierdo de la ecuación.
- Si se sabe h se puede hallar el lado derecho de la desigualdad fácilmente.

Dadas varias máquinas de aprendizaje, eligiendo un η fijo suficientemente pequeño, se elegirá la máquina que minimize el lado derecho de la ecuación.

2.1.2.1. Clasificación de puntos con hiperplanos orientados en \mathbb{R}^n

Si se tienen unos datos en \mathbb{R}^2 y el conjunto $f(\alpha)$ consiste en líneas rectas orientadas tal que dada una línea se pueden separar dos clases diferentes de puntos. De esta forma se pueden separar tres puntos pero no más. Es decir, la dimensión VC del conjunto de líneas orientadas en \mathbb{R}^2 es tres.

En \mathbb{R}^n , la complejidad de la máquina es $n+1$ ya que se pueden separar $n+1$ puntos como se quiera si se consideran hiperplanos en \mathbb{R}^n .

Teorema: Considere un conjunto de m puntos en \mathbb{R}^n , si se elige un punto cualquiera como origen, los m puntos pueden ser clasificados por hiperplanos orientados si y sólo si la posición de los vectores de los puntos restantes son linealmente independientes.

Corolario: La dimensión VC de un conjunto de hiperplanos en \mathbb{R}^n es $n+1$ ya que se pueden coger siempre $n+1$ puntos de los cuales uno será el origen y las posiciones de los vectores de los n puntos restantes son linealmente independientes.

2.1.2.2. Minimización del límite minimizando h

La confianza VC es una función que se incrementa monótonamente con h . Si el riesgo empírico es cero, se cogerá la máquina que tenga la dimensión VC mínima. Si no es

cero, se cogerá la máquina que minimice el término derecho (ecuación 2.4) [24] [25] [26].

Hay máquinas que con el mismo riesgo empírico y teniendo una dimensión VC mayor funcionan mejor.

2.1.2.3. Riesgo de minimización estructural (SRM)

El término de confianza VC depende de la clase de funciones elegidas. Mientras que el riesgo empírico y el riesgo real dependen de una función particular elegida al hacer el entrenamiento.

Se quiere encontrar el subconjunto dentro del conjunto de funciones elegidas tal que el límite del riesgo para ese subconjunto sea mínimo.

Se introduce una *estructura* dividiendo la clase entera de funciones en subconjuntos. Para cada subconjunto se debe de poder calcular h. La minimización del riesgo estructural (SRM) consiste en encontrar ese subconjunto de funciones que minimice el límite del riesgo actual (real). Para ello se entrenan una serie de máquinas, una para cada subconjunto, dentro del subconjunto el objetivo es minimizar el riesgo empírico. Se toma la máquina entrenada de la serie cuya suma de riesgo empírico y la confianza VC sea mínima.

2.1.3. Máquinas lineales de vectores soporte

Una máquina de vectores soporte (SVM) es un algoritmo que sigue el principio SRM (Structural Risk Minimization). En esta sección se desarrolla la SVM lineal. Ésta es generalizable al caso no lineal mediante el truco de los Núcleos [[32]]. En este trabajo no se utilizan las SVM no lineales, sin embargo se explican en el apéndice B. En este proyecto sólo se usa la regresión, la clasificación se explica por completitud.

2.1.3.1. Caso separable

Se etiquetan los datos de entrenamiento como $\{x_i, y_i\}$, donde $i=1, \dots, l$, $x_i \in \mathbb{R}^d$ e $y_i \in \{1, -1\}$. Se tiene un hiperplano que separa los ejemplos positivos de los negativos. Los puntos que están en el hiperplano cumplen la ecuación:

$$w \cdot x + b = 0 \quad (2.5)$$

donde b es cero si el hiperplano contiene al origen.

- w es normal al hiperplano y $\|w\|$ es la norma euclídea de w .
- $\frac{\|b\|}{\|w\|}$ es la distancia perpendicular del hiperplano al origen.
- $d_+(d_-)$ es la mínima distancia del hiperplano de separación a la muestra positiva(negativa) más cercana. El margen de separación del hiperplano es $d_+ + d_-$.

Para el caso en el que los datos son linealmente separables, el algoritmo de vectores soporte busca el hiperplano de separación con mayor margen, es decir maximizar la distancia d mostrada en la figura 2.1. Todos los datos de entrenamiento satisfacen:

$$x_i \cdot w + b \geq +1 \quad \text{para} \quad y_i = +1 \quad (2.6)$$

$$x_i \cdot w + b \leq -1 \quad \text{para} \quad y_i = -1 \quad (2.7)$$

Se pueden combinar estas dos desigualdades en:

$$y_i \cdot (x_i \cdot w + b) - 1 \geq 0 \quad \forall i \quad (2.8)$$

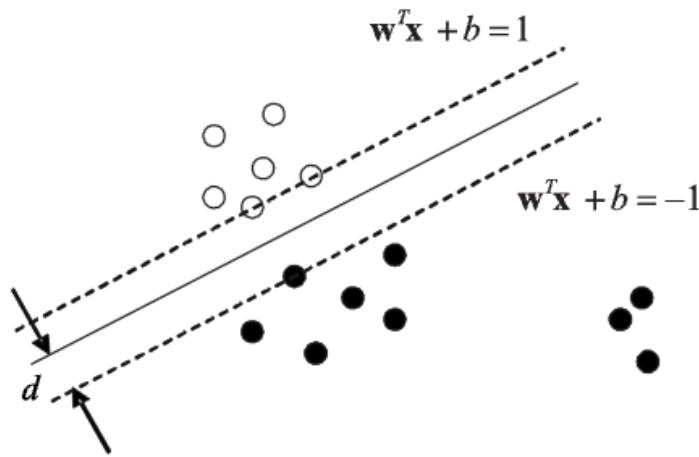


Figura 2.1: El procedimiento SVM consiste en poner el hiperplano tan lejos como sea posible de las muestras más cercanas.

Los puntos que están en el hiperplano H_1 cumplen: $x_i \cdot w + b = 1$, cuya distancia perpendicular al origen es: $\frac{|1-b|}{\|w\|}$.

Los puntos que están en el hiperplano H_2 cumplen: $x_i \cdot w + b = -1$, cuya distancia perpendicular al origen es: $\frac{|-1-b|}{\|w\|}$.

$d_+ = d_- = \frac{1}{\|w\|}$, el margen es igual a $\frac{2}{\|w\|}$.

H_1 y H_2 son paralelos (tienen la misma normal) y no hay puntos de entrenamiento entre ellos.

Minimizando $\|w\|^2$ y ajustándose a las restricciones de la desigualdad 2.8 se encuentran la pareja de hiperplanos que maximizan el margen.

A los puntos de entrenamiento que están en los hiperplanos H_1 o H_2 y que si se borrarán cambiarían la solución de los planos encontrados H_1 y H_2 (que son los que maximizan el margen) se les llama vectores soporte.

Se considera este problema con una formulación de Lagrange ya que los multiplicadores de Lagrange son más fáciles de tratar que las restricciones de la desigualdad 2.8 y los datos de entrenamiento sólo aparecerán como productos escalares entre vectores. Se introducen los multiplicadores de Lagrange: α_i , con $i=1, \dots, l$ de tal forma que:

$$L_P = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^l \alpha_i \quad (2.9)$$

Se debe minimizar L_P con respecto a \mathbf{w} y b y simultáneamente se requiere que se anulen las derivadas de L_P con respecto a todos los α_i , siendo siempre $\alpha_i \geq 0$. A este conjunto de restricciones se le llama C_1 . Esto es un problema de programación cuadrática convexa, ya que la función objetivo es convexa y todos los puntos que satisfacen las restricciones forman un conjunto convexo. Esto quiere decir que se puede solucionar el problema dual:

Maximizar L_P , sujeto a que se anule el gradiente de L_P con respecto a \mathbf{w} y b para valores de $\alpha_i \geq 0$. A este conjunto de restricciones se le llama C_2 . A esta formulación dual del problema también se le llama formulación de Wolfe.

El máximo de L_P sujeto a las restricciones de C_2 ocurre en los mismos valores de \mathbf{w} , b y α que el mínimo de L_P sujeto a las restricciones de C_1 .

Si se quiere que el gradiente de L_P se desvanezca con respecto a \mathbf{w} y b se tiene que cumplir:

$$\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \quad (2.10)$$

$$\sum_i \alpha_i y_i = 0 \quad (2.11)$$

Al sustituir estas ecuaciones en la ecuación 2.9:

$$L_D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \cdot \mathbf{x}_j \quad (2.12)$$

En notación matricial, se obtiene:

$$L_D = -\frac{1}{2} \boldsymbol{\alpha}^\top \mathbf{Y} \mathbf{X}^\top \mathbf{X} \mathbf{Y} \boldsymbol{\alpha} + \mathbf{1}^\top \boldsymbol{\alpha} \quad (2.13)$$

donde $\boldsymbol{\alpha}$ es el vector diagonal con los valores de α , $\mathbf{1}$ es un vector diagonal de unos, \mathbf{X} es una matriz con los vectores columna de \mathbf{x} e \mathbf{Y} es una matriz con los vectores columna de \mathbf{y} .

Las etiquetas de las fórmulas quieren decir: $P \rightarrow$ principal y $D \rightarrow$ dual. L_P y L_D surgen de la misma función objetivo pero con diferentes restricciones, la solución se encuentra minimizando L_P y maximizando L_D .

Hay un multiplicador de Lagrange α_i para cada punto de entrenamiento. Los vectores soporte tienen un $\alpha_i > 0$ y yacen en H_1 o en H_2 . El resto de los puntos tienen $\alpha_i = 0$ y yacen en H_1 o en H_2 o tienen $\alpha_i \neq 0$ y están en el lado de H_1 o en el de H_2 . Los vectores soporte son los elementos críticos del conjunto de entrenamiento.

Nótese que (2.13) es una forma cuadrática, por lo que tiene solución y ésta es única.

2.1.3.2. Las condiciones de Karush-Kuhn-Tucker

Juegan un papel central en la teoría y en la práctica de la optimización de restricciones. Las condiciones KKT son [26]:

$$\frac{\partial}{\partial w_v} L_P = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad v = 1, \dots, d \quad (2.14)$$

$$\frac{\partial}{\partial b} L_P = - \sum_i \alpha_i y_i = 0 \quad (2.15)$$

$$y_i \cdot (x_i \cdot w + b) - 1 \geq 0 \quad i = 1, \dots, l \quad (2.16)$$

$$\alpha_i \geq 0 \quad \forall i \quad (2.17)$$

$$\alpha_i (y_i \cdot (x_i \cdot w + b) - 1) = 0 \quad \forall i \quad (2.18)$$

Solucionar el problema de SVM es equivalente a encontrar la solución de las condiciones KKT. Aplicación: aunque w es explícitamente determinada por el procedimiento de entrenamiento, el umbral de b no. Usando las condiciones KKT, se puede hallar b fácilmente (ecuación 2.18) (eligiendo cualquier i para el cual $\alpha_i \neq 0$ se halla b , también se pueden tomar varias i y calcular la media de b).

2.1.3.3. Fase de test

Una vez la máquina de vectores soporte está entrenada, dado un dato de test x (entrada) se le asigna una etiqueta de clase dependiendo en qué lado del borde de decisión está el dato. La clase que le corresponde a x es $\text{sgn}(w \cdot x + b)$.

2.1.3.4. Caso no separable

Para tratar datos no separables hay que relajar las restricciones (ecuaciones 2.6 y 2.7) cuando sea necesario. Para ello se introducen unas variables positivas: ϵ_i , $i=1, \dots, l$, quedando las ecuaciones de la siguiente forma:

$$x_i \cdot w + b \geq +1 - \epsilon_i \quad \text{para} \quad y_i = +1 \quad (2.19)$$

$$x_i \cdot w + b \leq -1 + \epsilon_i \quad \text{para} \quad y_i = -1 \quad (2.20)$$

$$\epsilon_i \geq 0 \quad \forall i \quad (2.21)$$

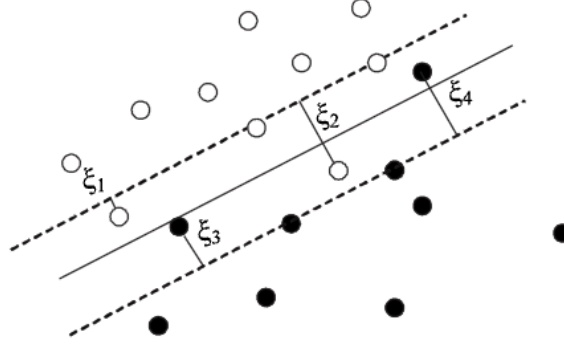


Figura 2.2: Hiperplanos de separación para el caso no separable.

- Si la muestra x_i está bien clasificada y fuera del margen, $\epsilon_i = 0$.
- Si la muestra x_i está bien clasificada pero dentro del margen, $0 < \epsilon_i < 1$.
- Si hay un error, es decir la muestra está mal clasificada, $\epsilon_i > 1$.

Por lo que $\sum_i \epsilon_i$ es un límite superior del número de errores en el entrenamiento. Se asigna un coste extra para los errores cambiando la función objetivo de $\frac{\|w\|^2}{2}$ a $\frac{\|w\|^2}{2} + C(\sum_i \epsilon_i)^k$, donde C , parámetro escogido por el usuario, es un compromiso entre la maximización del margen y la minimización de los errores y k es un entero positivo.

Si se elige $k=1$, ni ϵ_i ni los multiplicadores de Lagrange aparecen en el problema dual de Wolfe [27]. Las ecuaciones 2.10, 2.11 y 2.12 se siguen manteniendo igual con la única diferencia que α tiene un límite superior que es C . La ecuación principal de Lagrange:

$$L_P = \frac{1}{2} \|w\|^2 + C \sum_i \epsilon_i - \sum_i \alpha_i \{y_i(x_i \cdot w + b) - 1 + \epsilon_i\} - \sum_i \mu_i \epsilon_i \quad (2.22)$$

dónde μ_i son multiplicadores de Lagrange para asegurar que ϵ_i sea positiva.

Las condiciones KKT para el problema fundamental (principal) son:

$$\frac{\partial L_P}{\partial w_v} = w_v - \sum_i \alpha_i y_i x_{iv} = 0 \quad (2.23)$$

$$\frac{\partial L_P}{\partial b} = - \sum_i \alpha_i y_i = 0 \quad (2.24)$$

$$\frac{\partial L_P}{\partial \epsilon_i} = C - \alpha_i - \mu_i = 0 \quad (2.25)$$

$$y_i(x_i \cdot w + b) - 1 + \epsilon_i \geq 0 \quad (2.26)$$

$$\varepsilon_i \geq 0 \quad (2.27)$$

$$\alpha_i \geq 0 \quad (2.28)$$

$$\mu_i \geq 0 \quad (2.29)$$

$$\alpha_i \{y_i(x_i \cdot w + b) - 1 + \varepsilon_i\} = 0 \quad (2.30)$$

$$\mu_i \varepsilon_i = 0 \quad (2.31)$$

Donde $1 \leq i \leq$ número de puntos de entrenamiento y $1 \leq v \leq$ dimensión de los datos.

Como antes, se pueden utilizar las condiciones KKT (ecuaciones 2.30 y 2.31) para determinar el umbral b . De la ecuación 2.23 y 2.31 se puede sacar que $\varepsilon_i = 0$ si $\alpha_i < C$.

Al aplicar las condiciones de Karush-Kuhn-Tucker en la ecuación 2.22, se obtiene otra vez la ecuación 2.12 con las restricciones 2.28 y $\alpha_i \leq C$.

2.1.4. Máquinas de vectores soporte lineales para regresión (SVR)

2.4.1.1. Introducción

Un regresor lineal es una función $f(x) = w^T x + b$ que aproxima un conjunto de vectores $x \in \mathbb{R}^n$ a un conjunto de escalares $y \in \mathbb{R}$.

La regresión lineal ha sido tradicionalmente solucionada usando las aproximaciones de mínimos cuadrados. Es decir, la función de regresión es el hiperplano que ajusta los datos dados con el mínimo error cuadrático medio.

Sin embargo, el objetivo de SVR es diferente [28]: La idea principal es encontrar una función que ajuste los datos con una desviación menor que una cantidad ε para cada par x_i, y_i . Al mismo tiempo se requiere que la solución minimice $\|w\|$. Esto quiere decir que la SVR sólo minimizará los errores que sean iguales o mayores que ε . Así, con estas restricciones, se construye una máquina cuyos parámetros son una combinación lineal de aquellas muestras cuyo error es igual o mayor que ε .

2.4.1.2. Formulación

Si se ponen las ideas dichas en la introducción de forma matemática se tiene que:

$$L_P = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2.32)$$

Se trata de minimizar el error total más la norma de w .

Además, se tiene que cumplir:

$$y_i - w^T x_i - b \leq \xi_i + \varepsilon \quad (2.33)$$

$$-y_i + w^T x_i - b \leq \xi_i^* + \varepsilon \quad (2.34)$$

$$\xi_i, \xi_i^* \geq 0 \quad (2.35)$$

Lo que quiere decir que para cada muestra, los distintos casos posibles son:

- Si el error es positivo/(negativo) y su valor absoluto es mayor que ϵ , el error total será menor que $\xi_i + \epsilon / (\xi_i^* + \epsilon)$.
- Si el valor absoluto del error es menor que ϵ , ξ_i (ξ_i^*) serán cero.

Este concepto se puede observar en la figura 2.3

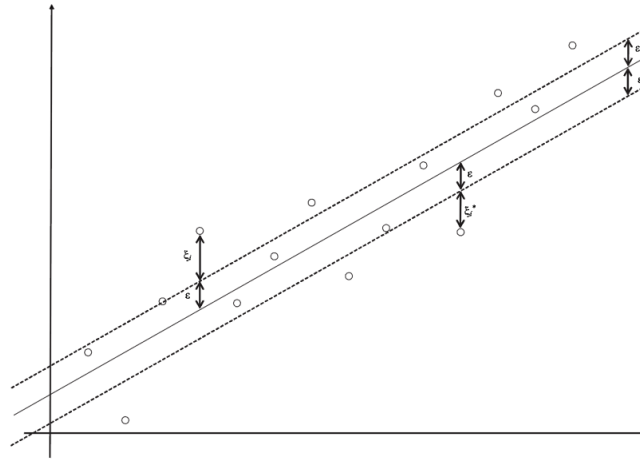


Figura 2.3: Únicamente las muestras que estén fuera del margen $\pm\epsilon$, serán parte de la solución

Se trata de minimizar la suma de las variables ξ_i (ξ_i^*). Únicamente aparecen las pérdidas de las muestras para las cuales el error es mayor que ϵ , por lo tanto la solución va a ser sólo función de esas muestras. La función de coste aplicada es lineal, de tal

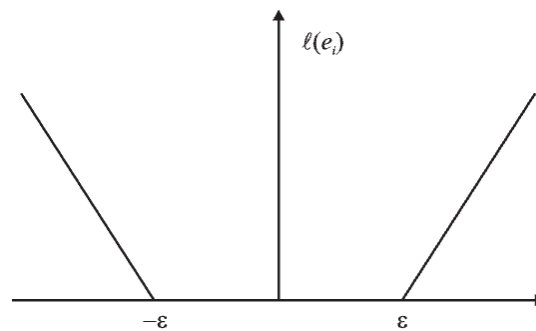


Figura 2.4: Vapnik o función de coste ϵ -insensible.

forma que el procedimiento es equivalente a la aplicación de la llamada Vapnik o función de coste ϵ -insensible, se muestra en la figura 2.4; Así la función de coste asociada a las pérdidas es:

$$\ell(e_i) = \begin{cases} 0 & |e_i| < \varepsilon \\ |e_i| - \varepsilon & |e_i| > \varepsilon \end{cases} \quad (2.36)$$

para $e = \xi_i + \varepsilon$, $e = -\xi_i^* - \varepsilon$.

En teoría se debería tener errores menores que ε y minimizar la norma de los parámetros. Sin embargo, en la práctica no se pueden tener todos los errores menores que ε . Para solucionar esto, se introducen unas nuevas variables: ξ_i y luego se minimizan.

Para solucionar el problema planteado en las ecuaciones 2.32, 2.33 y 2.34 se lleva a cabo la optimización de Lagrange, quedando la formulación dual de la siguiente manera:

$$L_d = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_{i'}) x_i^\top x_j (\alpha_i - \alpha_{j'}) + \sum_{i=1}^N ((\alpha_i - \alpha_{i'}) y_i - (\alpha_i + \alpha_{i'}) \varepsilon) \quad (2.37)$$

Sabiendo que:

$$0 \leq (\alpha_i - \alpha_{i'}) \leq C \quad (2.38)$$

El resultado importante de todo esto es la expresión del parámetro ω :

$$\omega = \sum_{i=1}^N (\alpha_i - \alpha_{i'}) x_i \quad (2.39)$$

y que:

$$\sum_{i=1}^N (\alpha_i - \alpha_{i'}) = 0 \quad (2.40)$$

Para hallar la desviación, b , es necesario encontrar las muestras que están dentro del margen. El error para esas muestras es ε y $\alpha_i, \alpha_{i'} < C$. Una vez que se identifican las muestras, se puede hallar b a partir de:

$$y_i - \omega^\top x_i - b + \varepsilon = 0 \quad (2.41)$$

$$-y_i + \omega^\top x_i + b + \varepsilon = 0 \quad (2.42)$$

para aquellas x_i para las cuales $\alpha_i, \alpha_{i'} < C$

En notación matricial, se tiene:

$$L_d = \frac{1}{2} (\alpha - \alpha')^\top \mathbf{X}^\top \mathbf{X} (\alpha - \alpha') + (\alpha - \alpha')^\top \mathbf{y} - \mathbf{1}^\top (\alpha - \alpha') \varepsilon \quad (2.43)$$

Nótese que, al igual que en (2.13), la ecuación 2.43 es una forma cuadrática, por lo que tiene solución y ésta es única.

2.1.5. Multirregresor SVM

2.1.6.1. Introducción Es una extensión de la máquina de vectores soporte (SVM) usada para regresión (SVR) para múltiples variables (M-SVR).

En vez de tener en cuenta cada componente individualmente, se usa una herramienta de regresión multidimensional, de tal forma que la estimación final es menos vulnerable al ruido. Tratando múltiples variables a la vez, se puede hacer una estimación más precisa de la salida disponiendo de pocos datos a la entrada de cada máquina de entrenamiento. Añadiendo la función de coste ϵ – *insensible*, definida ya en SVR, hará el sistema más robusto cuando aparezcan diferentes tipos de ruido y no linealidades.

Como se presentó en el apartado 2.4.1.2 de las máquinas lineales SVR, el problema de estimación de variables unidimensionales usando regresión (SVR) consiste en encontrar la correspondencia entre un vector de entrada $\mathbf{x} \in \mathbb{R}^d$ y su correspondiente salida $y \in \mathbb{R}$ dado un conjunto de muestras independientes e idénticamente distribuidas (i.i.d) $\{(x_i, y_i)\}$ donde $i = 0, \dots, n$.

En un caso general, SVR soluciona este problema encontrando el regresor \mathbf{w} y b que minimize: $\frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^N L_v(y_i - (\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b))$, donde $\phi(\cdot)$ es una transformación no lineal a un espacio de mayores dimensiones ($\phi(\cdot) \in \mathbb{R}^H$ y $H \geq d$). SVR puede ser solucionado únicamente con productos interiores entre $\phi(\cdot)$, es decir, no se necesita saber la correspondencia de las funciones no lineales, tan sólo hay que saber la función kernel $k(x_i, x_j) = \phi^T(x_i)\phi(x_j)$ que tiene que cumplir el teorema de Mercer (ver apéndice B). $L_v(\cdot)$ es conocida como la función de Vapnik o función de coste ϵ -insensible, figura 2.4, la cual es cero para $|y_i - (\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b)| < \epsilon$ e igual a $|y_i - (\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b)| - \epsilon$ para $|y_i - (\boldsymbol{\phi}^T(\mathbf{x}_i)\mathbf{w} + b)| \geq \epsilon$. La solución (\mathbf{w}, b) , está formada por una combinación lineal de las muestras de entrenamiento en el espacio transformado que tienen un error absoluto mayor o igual a ϵ .

En el caso en el que la salida sea un vector $\mathbf{y} \in \mathbb{R}^Q$, se necesita resolver un problema de regresión multidimensional en el que se tiene que encontrar un regresor \mathbf{w}^j y \mathbf{b}^j ($j = 1, \dots, Q$) para cada salida. Se puede generalizar la SVR unidimensional para solucionar el caso multidimensional, de tal forma que hay que minimizar:

$$L_P(\mathbf{W}, \mathbf{b}) = \frac{1}{2} \sum_{j=1}^Q \|\mathbf{w}^j\|^2 + C \sum_{i=1}^n L(u_i) \quad (2.44)$$

La función de Vapnik o función de coste ϵ -insensible puede ser extendida al caso multidimensional; con la norma L_1 , se necesitaría tener en cuenta cada dimensión de forma independiente, de tal forma que la complejidad de la solución aumentaría linealmente con el número de dimensiones. Si en vez de L_1 se usase L_2 , se consideraría una única restricción para todas las dimensiones, produciendo un único vector soporte para todas

las dimensiones. Se usaría:

$$\mathbf{L}(u) = \begin{cases} 0 & u < \varepsilon \\ u^2 - 2u\varepsilon + \varepsilon^2 & u \geq \varepsilon \end{cases} \quad (2.45)$$

Donde, $u_i = \|e_i\| = \sqrt{e_i^T e_i}$, sabiendo que $e_i^T = y_i^T - \phi^T(x_i)W - b^T$, $W = [w^1, \dots, w^Q]$, $b = [b^1, \dots, b^Q]^T$, y $\phi(\cdot)$ es una transformación no lineal a un espacio de mayor dimensión.

Para $\varepsilon = 0$ este problema se reduce a una regresión de mínimos cuadrados usando un kernel independiente para cada componente. Sin embargo, para $\varepsilon \neq 0$ la solución tendrá en cuenta todas las salidas para construir cada regresor y será capaz de obtener predicciones más robustas. El precio a pagar será que la resolución del problema no se puede hacer directamente, sino que será necesario un procedimiento iterativo. Se ha ideado una aproximación cuasi-Newtoniana en la cual cada iteración tiene como mucho la misma complejidad computacional que un procedimiento de mínimos cuadrados para cada componente. Este es un problema de mínimos cuadrados y el número de iteraciones necesarias para obtener el resultado final es pequeño; así, el procedimiento resultante es sólo un poco más complicado que el que se tenía cuando $\varepsilon = 0$. A este procedimiento se le llama IRWLS Iterative (Re-Weighted Least Square) [19].

2.2. Redes de Procesos Gaussianos

2.2.1. Introducción

Un Proceso Gaussiano es un proceso estocástico cuyas realizaciones consisten en valores aleatorios asociados a cada punto en un intervalo de tiempo (o de espacio), de forma que cada variable aleatoria tiene una distribución normal. Además, cada conjunto finito de estas variables aleatorias tiene una distribución normal multivariante.

Una red de Procesos Gaussianos es una interpretación bayesiana del algoritmo de mínimos cuadrados, donde se supone que el conjunto de respuestas del modelo a entrenar es un Proceso Gaussiano.

Las redes de Procesos Gaussianos pueden ser usadas en la regresión y clasificación. Mientras que las salidas para clasificación son etiquetas discretas, la regresión está relacionada con la predicción de cantidades continuas. En este capítulo describiremos los métodos de redes de Procesos Gaussianos para problemas de regresión.

2.2.2. Regresión

2.2.2.1. Espacio de pesos

El simple modelo de regresión lineal donde la salida es una combinación lineal de las entradas ha sido estudiado y utilizado extensamente. Sus principales virtudes son su sencillez en la aplicación e interpretación. Su principal inconveniente es que sólo permite una flexibilidad limitada; si la relación entre la entrada y la salida no puede ser aproximada por una función lineal, el modelo presentará unas predicciones pobres.

Tenemos un conjunto de entrenamiento D de n observaciones, $D = \{(x_i, y_i) | i = 1, \dots, n\}$, donde \mathbf{x} es un vector de entrada de dimensión D e y es una salida o etiqueta escalar (variable dependiente); el vector columna de las entradas para los n casos se agregan a la matrix de diseño X de dimensiones $D \times n$, y las etiquetas se recogen en el vector y , por lo que podemos escribir $D = (X, y)$. En la regresión las etiquetas son valores reales.

2.2.2.1.1. El Modelo Lineal Estándar

El análisis Bayesiano del modelo de regresión lineal estándar con ruido gaussiano es:

$$f(x) = x^T w, \quad y = f(x) + \varepsilon, \quad (2.46)$$

donde \mathbf{x} es el vector de entrada, \mathbf{w} es un vector de pesos del modelo lineal, f es el valor de la función e y es el valor de la etiqueta observada. Asumimos que los valores observados y difieren de los valores de la función $f(\mathbf{x})$ por un ruido auditivo y admitimos que este ruido sigue una distribución Gaussiana independiente e idénticamente distribuida con media cero y varianza σ_n^2 .

$$\varepsilon \sim N(0, \sigma_n^2). \quad (2.47)$$

Este ruido junto con el modelo da lugar a la *likelihood*, la densidad de probabilidad de las observaciones dados los parámetros, los cuales están factorizados sobre los casos en el conjunto de entrenamiento, dando lugar a:

$$\begin{aligned} p(y|X, w) &= \prod_{i=1}^n p(y_i|x_i, w) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(y_i - x_i^T w)^2}{2\sigma_n^2}\right) \\ &= \frac{1}{(2\pi\sigma_n^2)^{n/2}} \exp\left(-\frac{1}{2\sigma_n^2} |y - X^T w|^2\right) = N(X^T w, \sigma_n^2 I), \end{aligned} \quad (2.48)$$

donde $|z|$ es la longitud euclidiana del vector z . En el formalismo Bayesiano necesitamos especificar un previo (*prior*) acerca de los parámetros, expresando nuestras creencias sobre ellos antes de que veamos las observaciones. Aproximamos el vector de pesos previo como una gaussiana de media cero con matriz de covarianzas Σ_p :

$$w \sim N(0, \Sigma_p). \quad (2.49)$$

La inferencia en el modelo lineal bayesiano se basa en la distribución posterior de los pesos, calculado por la regla de Bayes:

$$p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)}, \quad (2.50)$$

donde la constante de normalización también conocida como la probabilidad marginal, es independiente de los pesos y es dada por:

$$p(y|X) = \int p(y|X, w)p(w)dw. \quad (2.51)$$

Reconocemos la forma de la distribución posterior como una Gaussiana de media \bar{w} y matriz de covarianza A^{-1}

$$p(w|X, y) \sim N(\bar{w} = \frac{1}{\sigma_n^2} A^{-1} Xy, A^{-1}) \quad (2.52)$$

donde $A = \sigma_n^{-2} X X^T + \Sigma_p^{-1}$. Hay que tener en cuenta que para este modelo, la media de la distribución posterior $p(w|X, y)$ es también su moda, que además se conoce como el máximo a posteriori (MAP) de la estimación de w .

Para hacer predicciones para un caso de test, promediamos todos los posibles valores de los parámetros ponderados por su probabilidad posterior. De esta forma, la distribución predictiva para $f_* = f(x_*)$, con x_* obtenido por un promedio de la salida de todos los posibles modelos lineales es:

$$\begin{aligned} p(f_*|x_*, X, y) &= \int p(f_*|x_*, w)p(w|X, y)dw \\ &= N(\frac{1}{\sigma_n^2} x_*^T A^{-1} Xy, x_*^T A^{-1} x_*). \end{aligned} \quad (2.53)$$

La distribución predictiva es otra vez Gaussiana con media dada por la media posterior de los pesos multiplicada por la entrada de test (x_*^T). La varianza de la distribución predictiva es una forma cuadrática de la entrada de test con la matriz de covarianzas posterior.

En este proyecto se utiliza la regresión lineal. La explicación de la regresión no lineal se encuentra en el apéndice C.

2.2.2.2. Espacio de funciones (Function-space)

Una manera alternativa y equivalente de llegar a resultados idénticos a la sección anterior es considerando la inferencia directamente en el espacio de funciones. Utilizamos un proceso Gaussiano (GP) para describir una distribución sobre funciones. Así:

Definición: Un proceso Gaussiano es una colección de variables aleatorias, un número finito de las cuales tienen una distribución Gaussiana conjunta.

Un proceso Gaussiano se define por su función de media y su función de covarianza. Definimos la función de media $m(x)$ y la función de covarianza $k(x, x')$ de un proceso real $f(x)$ como:

$$\begin{aligned} m(x) &= E[f(x)], \\ k(x, x') &= E[f(x) - m(x)][f(x') - m(x')], \end{aligned} \quad (2.54)$$

por tanto, escribimos el proceso Gaussiano como:

$$f(x) \sim GP(m(x), k(x, x')). \quad (2.55)$$

En nuestro caso las variables aleatorias representan el valor de la función $f(x)$ en la posición x . A veces, los procesos Gaussianos se definen a través del tiempo, por ejemplo cuando el índice de las variables aleatorias es el tiempo. Este no es el caso en nuestro uso de GPs, donde X es el conjunto de posibles entradas. Por conveniencia utilizamos la enumeración de los casos en el conjunto de entrenamiento para identificar las variables aleatorias tal que $f_i = f(x_i)$ es la variable aleatoria correspondiente al caso (x_i, y_i) .

Un simple ejemplo de proceso Gaussiano puede ser obtenido de nuestro modelo de regresión lineal Bayesiano $f(x) = \phi(x)^T w$ con $w \sim N(0, \Sigma_p)$. Por tanto la media y covarianza son:

$$\begin{aligned} E[f(x)] &= \phi(x)^T E[w] = 0, \\ E[f(x)f(x')] &= \phi(x)^T E[ww^T] \phi(x') = \phi(x)^T \Sigma_p \phi(x'). \end{aligned} \quad (2.56)$$

Así $f(x)$ y $f(x')$ son conjuntamente Gaussianas con media cero y covarianza dada por $\phi(x)^T \Sigma_p \phi(x')$. De hecho, los valores de la función $f(x_1), \dots, f(x_n)$ correspondiente a cualquier número de puntos de entrada n son conjuntamente Gaussianas.

La función de covarianza será la exponencial cuadrada (*squared exponential*). La función de covarianza especifica la covarianza entre pares de variables aleatorias:

$$\text{cov}(f(x_p), f(x_q)) = k(x_p, x_q) = \exp\left(-\frac{1}{2}|x_p - x_q|^2\right). \quad (2.57)$$

La covarianza entre las salidas se escribe como una función de las entradas. Para esta particular función, la covarianza es casi uno entre variables cuyas entradas están muy próximas, y disminuye a medida que su distancia en el espacio de entrada aumenta.

La especificación de la función de covarianza implica una distribución sobre funciones. Podemos dibujar muestras de la distribución de funciones evaluadas en cualquier número de puntos; elegimos un número de puntos, X_* y escribimos la matriz de covarianzas correspondiente usando la ecuación 2.57 elemento a elemento. Por tanto generamos un vector aleatorio Gaussiano con esta matriz de covarianzas:

$$f_* \sim N(0, K(X_*, X_*)), \quad (2.58)$$

Predicción usando observaciones ruidosas

Suponemos ruido aditivo Gaussiano independiente e idénticamente distribuido ϵ con varianza σ_n^2 , la función de covarianza cumple:

$$\text{cov}(y_p, y_q) = k(x_p, x_q) + \sigma_n^2 \delta_{pq} \quad o \quad \text{cov}(y) = K(X, X) + \sigma_n^2 I, \quad (2.59)$$

donde δ_{pq} es la delta Kronecker que vale 1 si $p=q$ y 0 en cualquier otro caso.

La distribución predictiva para regresión del proceso Gaussiano es:

$$\begin{aligned} f_* | X, y, X_* &\sim N(\bar{f}_*, \text{cov}(f_*)), \quad \text{donde} \\ \bar{f}_* &= E[f_* | X, y, X_*] = K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} y, \\ \text{cov}(f_*) &= K(X_*, X_*) - K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*). \end{aligned} \quad (2.60)$$

Utilizamos la notación $K = K(X, X)$ y $K_* = K(X, X_*)$. En el caso de que haya un único punto de test x_* , escribimos $k(x_*) = k_*$ para indicar el vector de covarianzas entre los puntos de test y los 'n' puntos de entrenamiento. Por tanto la distribución predictiva se reduce a:

$$\begin{aligned} \bar{f}_* &= K_*^T (K + \sigma_n^2 I)^{-1} y, \\ V[f_*] &= k(x_*, k_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_*. \end{aligned} \quad (2.61)$$

La media de la distribución predictiva es una combinación lineal de las observaciones y ; esto se refiere a veces como *un predictor lineal*. Otra forma de observar esta ecuación es verla como una combinación lineal de n funciones kernel, cada una centrada en un punto de entrenamiento:

$$\bar{f}(x_*) = \sum_{i=1}^n \sigma_i k(x_i, x_*) \quad (2.62)$$

donde $\sigma = (K + \sigma_n^2 I)^{-1} y$.

Introducimos la probabilidad marginal $p(y|X)$. Esta probabilidad se define:

$$p(y|X) = \int p(y|f, X) p(f|X) df. \quad (2.63)$$

El término probabilidad *marginal* se refiere a la marginación en los valores de la función \mathbf{f} . Bajo el modelo del proceso Gaussiano, $f|X \sim N(0, K)$, o:

$$\log p(f|X) = -\frac{1}{2} f^T K^{-1} f - \frac{1}{2} \log |K| - \frac{n}{2} \log 2\pi, \quad (2.64)$$

y la probabilidad es una Gaussiana factorizada $y|f \sim N(f, \sigma_n^2 I)$, por lo que podemos calcular la integral realizando el logaritmo de la probabilidad marginal:

$$\log p(y|X) = -\frac{1}{2}y^T(K + \sigma_n^2 I)^{-1}y - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2}\log 2\pi. \quad (2.65)$$

Este resultado puede obtenerse directamente observando que $y \sim N(0, K + \sigma_n^2 I)$.

2.2.3. Cálculo de la matriz de covarianzas

La matriz de covarianzas es A^{-1} , donde:

$$\begin{aligned} A &= \sigma_n^{-2}XX^T + \Sigma_p^{-1} = \sigma_n^{-2}(XX^T + \sigma_n^2 I) = \frac{1}{\sigma_n^2}(XX^T + \sigma_n^2 I) \\ &= \sigma_n^{-2}Q\Lambda Q^T + I = \sigma_n^{-2}Q(\Lambda + \sigma_n^2 I)Q^T \end{aligned} \quad (2.66)$$

y donde σ_n^{-2} es la varianza de los datos, X son los datos y $Q\Lambda Q^T$ es la descomposición de XX^T en componentes principales.

Por lo tanto, la matriz de covarianzas A^{-1} se define como:

$$A^{-1} = \sigma_n^2 Q(\Lambda + \sigma_n^2 I)^{-1} Q^T \quad (2.67)$$

siendo $(\Lambda + \sigma_n^2 I)^{-1}$ una matriz diagonal cuyos elementos son $\frac{1}{\lambda_i + \sigma_n^2}$.

Sabiendo que $K = X^T X$ y $R = XX^T$, los autovalores de K son iguales a los autovalores de R no nulos. También sucede que los autovectores de R son $q_i = \frac{1}{\sqrt{\lambda_i}}XV_i$, donde V_i son los autovectores de K .

Necesitamos un elemento de la matriz de covarianzas, por tanto:

$$\begin{aligned} a_{ij} &= \sigma_n^2 [q_1(i) \dots q_M(i)] [\Lambda + \sigma_n^2 I]^{-1} [q_1(j) \dots q_M(j)]^T \\ &= \sigma_n^2 [q_1(i) \dots q_M(i)] \left[\frac{1}{\lambda_1 + \sigma_n^2} q_1(j) \dots \frac{1}{\lambda_M + \sigma_n^2} q_M(j) \right]^T \end{aligned} \quad (2.68)$$

Descomponiendo a_{ij} nos queda:

$$a_{ij} = \sigma_n^2 \sum_{k=1}^N q_k(i) q_k(j) \frac{1}{\lambda_k + \sigma_n^2} \quad (2.69)$$

donde $q_k(i) = \frac{1}{\sqrt{\lambda_k}} X_i V_k$ corresponde a la fila i de X .

$$\begin{aligned}
a_{ij} &= \sigma_n^2 \sum_{k=1}^N X_i V_k \frac{1}{\lambda_k + \sigma_n^2} \frac{1}{\lambda_k} V_k^T X_j^T \\
&= \sigma_n^2 X_i \sum_{k=1}^N V_k V_k^T \frac{1}{\lambda_k + \sigma_n^2} \frac{1}{\lambda_k} X_j^T \\
&= \sigma_n^2 X_i V \Lambda^{-2} V^T X_j^T \\
&= X_i K^{-2} X_j^T
\end{aligned} \tag{2.70}$$

Por consiguiente, el elemento de la matriz de covarianzas que necesitamos quedaría:

$$a_{ij} = X_i (X^T X)^{-2} X_j^T \tag{2.71}$$

Estas varianzas se usarán para el cálculo de estimaciones tipo T de Student β/σ tanto para el Modelo Lineal General como para β calculado con Redes de Procesos Gaussianos o Máquinas de Vectores Soporte.

Todo esto se hace porque sólo necesitamos la diagonal de la matriz de covarianzas o una de sus filas. Además la dimensión de dicha matriz es extremadamente grande, lo que conlleva un cálculo computacional enorme.

La matriz de covarianzas puede ser lineal y no lineal:

- **Matriz de covarianza lineal:** La función Kernel o matriz de covarianzas se calcula como el producto de los datos transpuestos por los datos sin transponer, es decir $K = X'X$, y a partir de dicha matriz obtener la media predictiva (β) y la varianza para Redes de Procesos Gaussianos, así como la media para GLM y SVM.
- **Matriz de covarianza no lineal:** En este caso, utilizamos una función kernel de base radial gaussiana. Una vez obtenida la matriz de covarianzas, calculamos la media y la varianza para Redes de Procesos Gaussianos, así como la media obtenida con GLM y SVM.

2.2.4. Sumario

Las máquinas de vectores soporte son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik. De manera más formal, una SVM construye un hiperplano o conjunto de hiperplanos en un espacio de dimensionalidad muy alta (o incluso infinita) que puede ser utilizado en problemas de clasificación o regresión. Una buena separación entre las clases permitirá una clasificación correcta.

Dado un conjunto de datos de entrenamiento $\{x_i, y_i\}$, se trata de entrenar unas máquinas de vectores soporte (SVM) mediante las cuales poder clasificar nuevas entradas. Para ello tiene que haber un compromiso entre el número de muestras a entrenar y la complejidad

de las máquinas. Se trata de encontrar la máquina que minimice el límite del riesgo. Las máquinas de vectores soporte lineales a partir de las muestras de entrenamiento, definen un par de hiperplanos (mediante unos puntos característicos, vectores soporte); los nuevos datos pertenecerán a una clase u otra dependiendo del hiperplano en el que se encuentren. En el caso no separable, se introducirá un margen de error para poder realizar la clasificación. En el caso de las máquinas de vectores soporte para regresión (SVR), se trata de ajustar todos los datos al par de hiperplanos definido; se minimizarán los errores que sean mayores o iguales que la mitad del margen (ϵ). En el caso no lineal, tanto para SVM's como SVR's, se define un kernel mediante el cual se hace una transformación de los datos a un espacio de dimensión mayor (puede ser infinita) logrando así una separación lineal en un espacio diferente; una vez hecho esto, todo lo definido para el caso lineal puede ser aplicado. Para el caso del multirregresor SVM, en el que la salida es un vector y , se generaliza la SVR unidimensional a través de un algoritmo iterativo.

Las SVM tienen una sólida base matemática. Además tienen una elevada capacidad de generalización, comparado con otras redes neuronales y no tienen demasiados parámetros libres, comparadas con otras estrategias.

Por otro lado, un Proceso Gaussiano es un proceso estocástico cuyas realizaciones consisten en valores aleatorios asociados a cada punto en un intervalo de tiempo, de forma que cada variable aleatoria tiene una distribución normal. Además, cada conjunto finito de estas variables aleatorias tiene una distribución normal multivariante.

Una red de Procesos Gaussianos es una interpretación bayesiana del algoritmo de mínimos cuadrados, donde se supone que el conjunto de respuestas del modelo a entrenar es un Proceso Gaussiano.

DESCRIPCIÓN DE LOS ALGORITMOS, DATOS Y RESULTADOS

3.1. Descripción de los algoritmos

3.1.1. Modelo de datos

El objetivo de este proyecto es programar y probar una serie de algoritmos para su posterior evaluación, tomando estímulos motores del cerebro humano. Desarrollamos un algoritmo univariante denominado Modelo Lineal General (GLM) y dos algoritmos multivariantes conocidos como Proceso Gaussiano (GP) y Máquinas de Vectores Soporte (SVM).

Dichos algoritmos se han probado con datos de entrada basados en fMRI. En un experimento fMRI se obtiene una serie temporal de N imágenes tridimensionales que indican la variación de la oxigenación en el cerebro. Estas imágenes se almacenan a efectos de su representación visual en matrices tridimensionales con una ordenación siguiendo los tres ejes del espacio de la imagen. Sin embargo, a efectos de procesamiento de la señal, los valores de estas matrices (voxels) se redistribuyen de forma arbitraria en vectores $x[n]$, donde $0 \leq n \leq N - 1$ es el índice temporal de la serie. La concatenación en columna de todos estos vectores, cada uno de los cuales se representa en formato fila, constituye la matriz de respuesta hemodinámica X , de dimensiones $N \times M$.

3.1.2. Modelo de señal

El experimento fMRI consiste en la realización de una actividad (motora, cognitiva...) a intervalos preestablecidos que se señalan mediante algún medio sensorial al sujeto. Las señales indicadoras binarias denominadas vectores de referencia representan el inicio y final de los estímulos y actividades en cada bloque. Estas señales se filtran mediante un filtro gaussiano que aproxima la respuesta hemodinámica del cerebro a la actividad en cuestión.

CAPÍTULO 3. DESCRIPCIÓN DE LOS ALGORITMOS, DATOS Y RESULTADOS

El conjunto de los vectores de referencia filtrados $y_l[n]$, $0 \leq l \leq L - 1$ constituye la matriz de diseño \mathbf{Y} , de dimensiones $N \times L$.

3.1.3. Modelo del algoritmo univariante: GLM

Este modelo se diseña de forma que, para cada voxel, se estima linealmente la respuesta hemodinámica a partir de la matriz de diseño. Este método es univariable, es decir, la estimación de cada vector β_{m1} sólo depende de la respuesta hemodinámica en el voxel m . En notación matricial puede ser escrito de la siguiente forma:

$$\mathbf{X} = \mathbf{Y}\beta + \mathbf{E} \quad (3.1)$$

donde \mathbf{X} es la respuesta cerebral e \mathbf{Y} es la matriz de diseño. El conjunto de parámetros β es una matriz de dimensiones $L+1 \times M$. Las L primeras filas representan una imagen del cerebro cuyos voxels indican el nivel de presencia de cada uno de los estímulos en la correspondiente zona del cerebro. La última fila representa el sesgo de cada uno de los voxels en el estimador, y en principio, carece de interés clínico. Y la matriz \mathbf{E} contiene el error de estimación por cada voxel e instante de tiempo.

En el caso de GLM, si la inversa de $\mathbf{Y}^T \mathbf{Y}$ existe, lo que ocurre cuando la matriz de diseño es de rango completo, las estimaciones de mínimos cuadrados se pueden calcular de la siguiente manera:

$$\hat{\beta}_{GLM} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{X} \quad (3.2)$$

3.1.4. Modelo de los algoritmos multivariantes: GP y SVM

En el caso de GP, el vector de pesos se puede calcular:

$$p(w|X, y) \sim N(\bar{w} = \frac{1}{\sigma_n^2} A^{-1} Xy, A^{-1}) \quad (3.3)$$

donde $A = \sigma_n^{-2} X X^T + \Sigma_p^{-1}$. De esta forma se utiliza una matriz de covarianzas de dimensión $M \times M$ no tratable, por tanto se usa una transformación a dual que utiliza $X^T X$ de dimensión $N \times N$, que es lo que se define como matriz de covarianzas o Kernel (K). w es el vector de pesos, denominado también β e indica la activación cerebral en cada uno de los voxels.

Por último, en el caso de SVM, es necesario una buena elección del kernel. Para ello hay que elegir una serie de parámetros como son C (compromiso entre el error y el coste computacional) y ϵ (epsilon). Los errores menores que ϵ no se tendrán en cuenta. En este proyecto hemos elegido $C=100$ y $\epsilon = 0$ para coger la mayor cantidad de información posible, ya que a mayor ϵ tomamos menos vectores de pesos. Tenemos que definir las opciones para nuestro modelo SVM y calculamos β para el caso dual.

También variamos el parámetro ϵ para ver las diferencias existentes y comprobar que a menor ϵ los mapas de activación se observan menos suavizados y con más artefactos.

Todo esto se realiza utilizando una matriz de covarianzas lineal $K = X^T X$. Además en este proyecto también se usa una matriz de covarianzas no lineal $K_{ij} = \exp(-\frac{\|X_i - X_j\|^2}{2\sigma^2})$ calculando nuevamente β en el caso de los algoritmos multivariantes.

Para calcular los mapas estadísticos se calculará σ_{ii} para cada voxel i usando el método de la sección 2.2.3.

3.1.5. Descripción del problema

Se trata de programar y probar una serie de algoritmos tanto univariantes utilizando GLM como multivariantes usando Proceso Gaussiano y SVM con el fin de obtener mapas de activación β , utilizando para ello estímulos motores del cerebro humano. Esto se desarrollará empleando para un caso la matriz de covarianzas lineal y para otro caso la matriz de covarianzas no lineal, comparando los resultados obtenidos.

Se ha implementado un algoritmo que a partir de unos datos de entrada consistentes en secuencias de imágenes temporales obtenidas mediante fMRI sea capaz de indicar de forma precisa las zonas del cerebro que son importantes en la actividad motora.

Las áreas funcionales del cerebro humano son: médula espinal, cerebelo, paretal, temporal, occipital, subcortical y frontal. Para el lado derecho como para el izquierdo (figura 3.1).

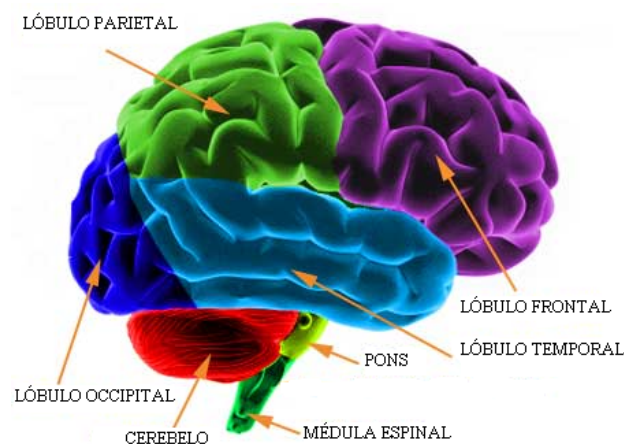


Figura 3.1: Áreas del cerebro.

CAPÍTULO 3. DESCRIPCIÓN DE LOS ALGORITMOS, DATOS Y RESULTADOS

La corteza cerebral incluye la corteza motora, la corteza sensorial y partes de la corteza vinculadas con la visión, la audición y el habla. En las cortezas motora y sensorial, los dos hemisferios cerebrales son imágenes especulares uno del otro: el hemisferio derecho controla y recibe información del lado izquierdo del cuerpo, y viceversa. Dichas cortezas (motora y sensorial) rodean al cerebro como auriculares. En la siguiente imagen 3.2 se puede ver las distintas partes de la corteza cerebral humana.

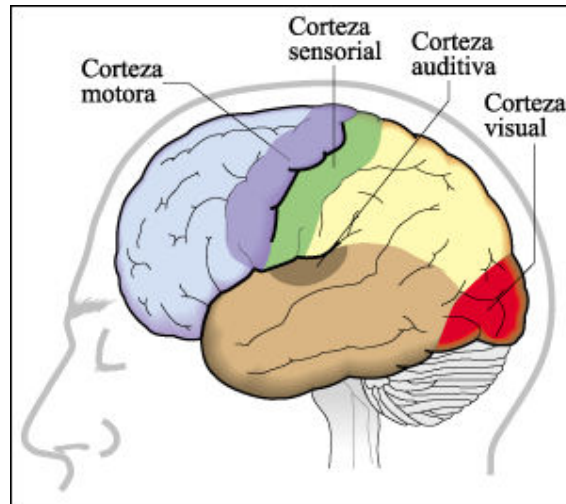


Figura 3.2: Corteza cerebral humana.

El área del lóbulo frontal inmediatamente anterior al surco central contiene las neuronas relacionadas con la integración de actividades llevadas a cabo por los músculos estriados esqueléticos: la corteza motora. En nuestro proyecto se verá dicho cortex motor activado debido a la tarea que los sujetos debían de realizar.

Inmediatamente detrás del surco central, en el lóbulo parietal, se encuentra la denominada corteza sensorial. Está relacionada con la recepción de estímulos táctiles (tacto), así como de estímulos vinculados al gusto, la temperatura y el dolor.

En el lóbulo temporal, parcialmente enterrado en el surco lateral, se encuentra la corteza auditiva que constituye el centro de procesamiento de las señales enviadas por las neuronas sensoriales del oído.

En definitiva, en este proyecto sólo nos interesa el área del lóbulo frontal debido a que los sujetos tenían como deber una tarea motora.

3.2. Datos

Sujetos y experimentos

Diez sujetos sanos (todos diestros) fueron estudiados usando un escáner Siemens de 3.0 Teslas. Los estímulos fueron presentados mediante unos auriculares compatibles con la máquina de Resonancia Magnética. El paradigma consistía en un bloque de 20 segundos de actividad y 20 segundos de reposo. A los sujetos se les pidió que tocasen con un dedo alternativo (con extensión máxima del dedo) un panel con botones de respuesta al ritmo de un tono auditivo.

Adquisición de datos

Los datos fMRI fueron adquiridos usando una imagen echo-planar con tiempo de repetición (TR) de 2000 milisegundos, tiempo de echo (TE) de 30 milisegundos, un ángulo de inclinación de 90° , un tamaño de matriz de 64×64 píxeles y 37 cortes. Los cortes son de 3 mm de espesor con un 25 % de desfase. Las dimensiones del voxel son de $3 \times 3 \times 3.75$ mm. Se recogieron 160 volúmenes para un tiempo total de medición de 320 segundos por sesión de escaneamiento. El conjunto de datos utilizado en los siguientes experimentos consiste de 10 estudios de 10 sujetos distintos.

3.3. Sumario

El esquema de los pasos desarrollados en este proyecto es el siguiente:

Se parte del conjunto de secuencias de imágenes temporales obtenidas a partir de las fMRI's realizadas a los distintos sujetos. Son los datos de entrada de los algoritmos.

Enmascaramos los datos y cogemos aquellos píxeles que sean distintos de cero, es decir aquellos que tengan información.

Calculamos la función Kernel o matriz de covarianza, que será distinta para el caso lineal y no lineal.

Calculamos el vector de pesos (β) utilizando el Modelo Lineal General (GLM).

Obtenemos β y la varianza usando Proceso Gaussiano (GP).

Implementamos el vector de pesos utilizando Máquinas de Vectores Soporte (SVM).

Dibujamos el mapeo inverso de β y $\frac{\beta}{\sqrt{G^2}}$ usando GLM.

CAPÍTULO 3. DESCRIPCIÓN DE LOS ALGORITMOS, DATOS Y RESULTADOS

Representamos el mapeo inverso de β y $\frac{\beta}{\sqrt{\sigma^2}}$ obtenido mediante GP y SVM.

Diagrama de bloques

Para tener una visión global y más clara de los algoritmos desarrollados se muestran los siguientes diagramas de bloques en los cuales se observa la secuencia de pasos llevada a cabo.

A continuación se muestra el diagrama de bloques del algoritmo univariante GLM.

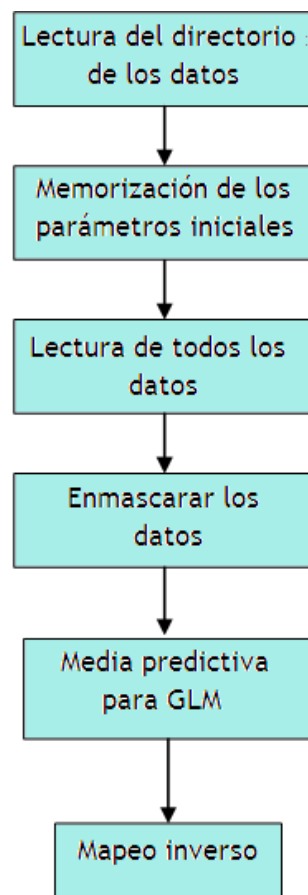


Figura 3.3: Diagrama de bloques del algoritmo univariante (GLM).

A continuación se puede observar el diagrama de bloques de los algoritmos multivariantes desarrollados en este proyecto: GP y SVM.

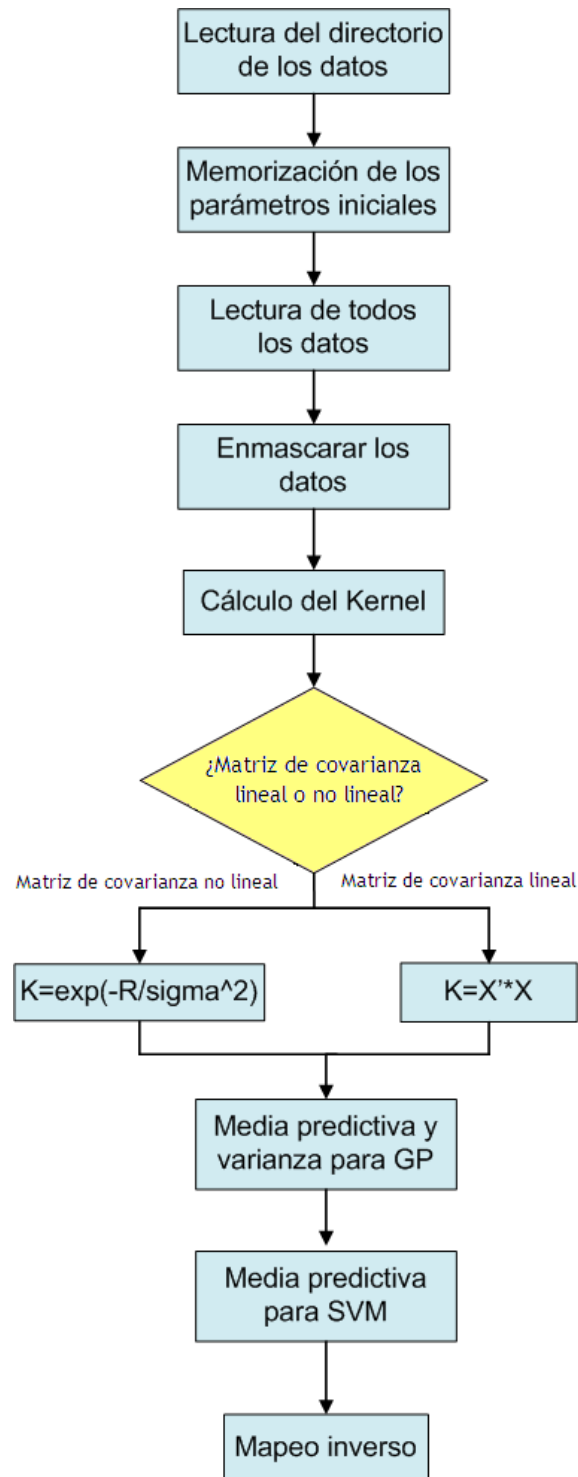


Figura 3.4: Diagrama de bloques de los algoritmos multivariantes (GP y SVM).

3.4. Resultados

Un mapa de activación es un mapa en el cual se puede ver la activación de cada una de las áreas funcionales. El rojo indicará mayor activación en un área del cerebro determinada. Este mapa de activación se obtiene a partir del mapeo inverso del vector de pesos β .

Se van a representar cinco resultados para el caso de utilizar la matriz de covarianza lineal y otros cinco resultados para el caso de la matriz de covarianza no lineal, que son:

Mapa de activación del vector de pesos (β) empleando GLM.

Mapa de activación del vector de pesos (β) usando GP.

Mapa de activación de β utilizando SVM.

Mapa de activación del cociente entre el vector de pesos y la varianza ($\beta/\text{varianza}$) con GLM.

Mapa de activación del cociente entre el vector de pesos y la varianza ($\beta/\text{varianza}$) empleando GP.

Mapa de activación del cociente entre $\beta/\text{varianza}$ con SVM.

Mapa de activación de β con SVM variando el parámetro ϵ .

Mapa de activación del cociente entre $\beta/\text{varianza}$ con SVM variando el parámetro ϵ .

Comenzamos viendo los resultados en el caso de utilizar una matriz de covarianzas lineal.

3.4.1. Caso 1: Matriz de covarianzas lineal

A continuación se muestran los resultados obtenidos en cuatro estudios para el caso de utilizar la matriz de covarianzas lineal. Hemos elegido aleatoriamente cuatro de los diez estudios realizados.

ESTUDIO 1

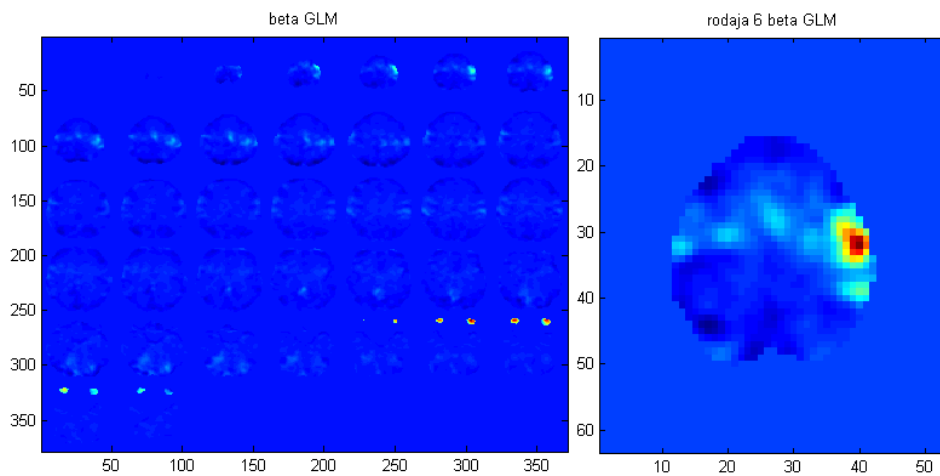


Figura 3.5: Mapa de activación de β usando *GLM* del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

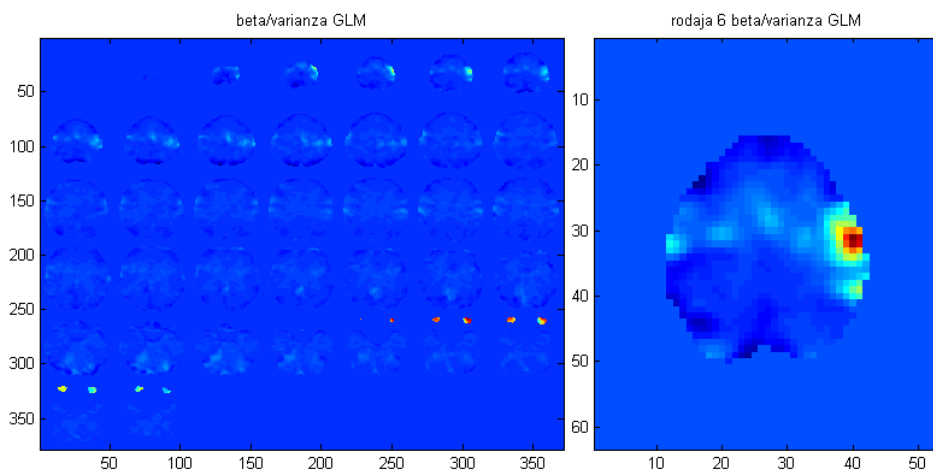


Figura 3.6: Mapa de activación del cociente entre el vector de pesos y la varianza usando *GLM* del estudio 1. A la derecha la rodaja con mayor activación cerebral.

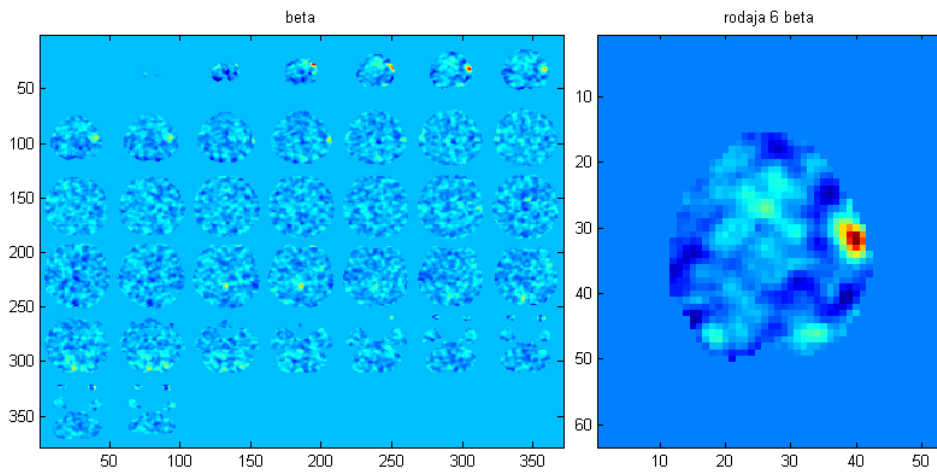


Figura 3.7: Mapa de activación de β usando **GP** del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

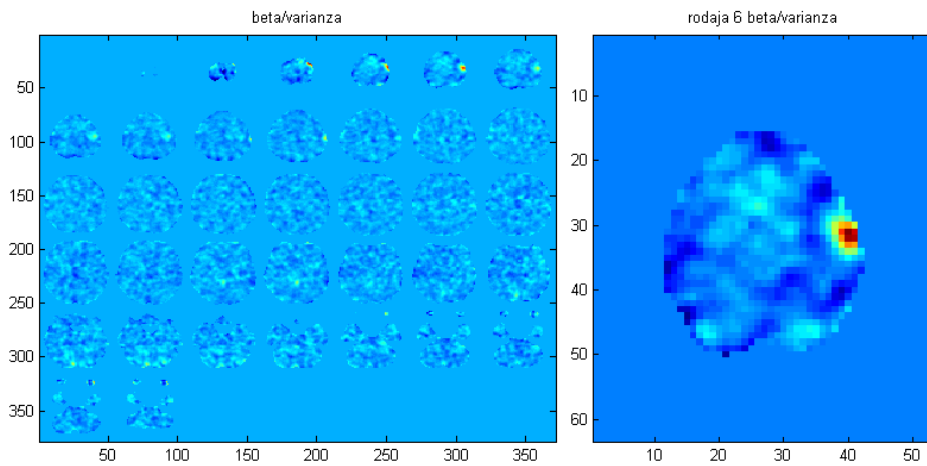


Figura 3.8: Mapa de activación del cociente entre el vector de pesos y la varianza usando **GP** del estudio 1. A la derecha la rodaja con mayor activación cerebral.

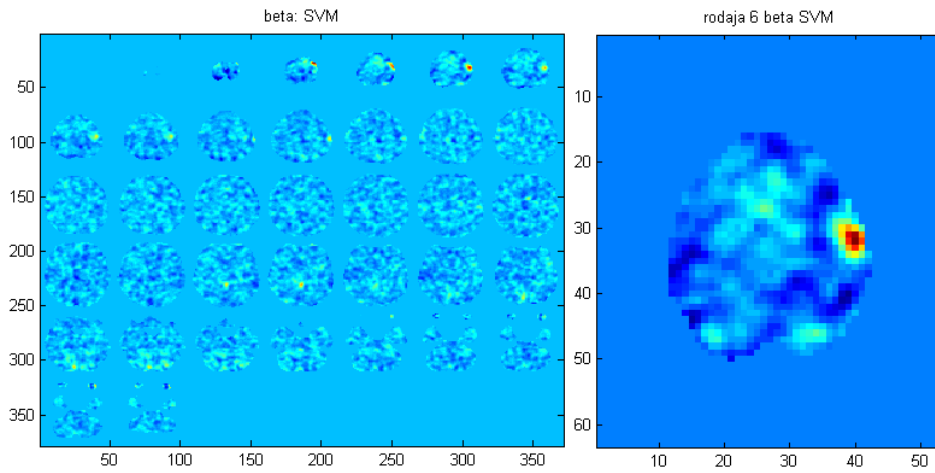


Figura 3.9: Mapa de activación de β usando SVM del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

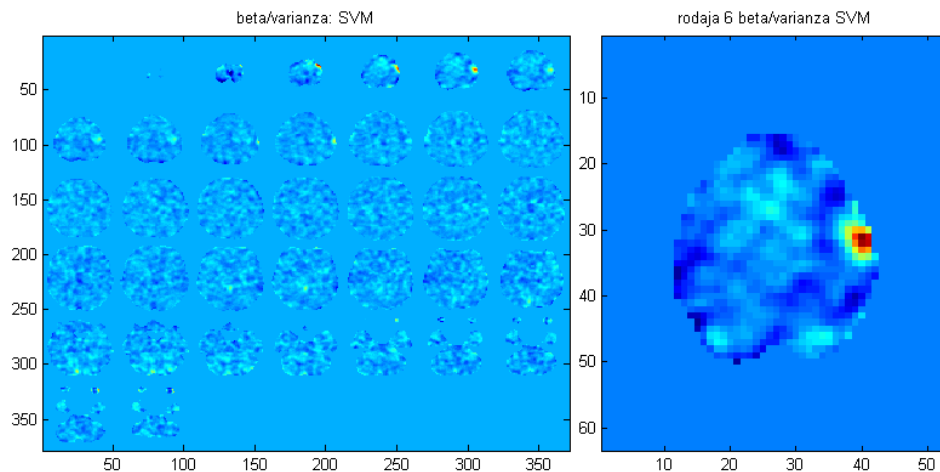


Figura 3.10: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 1. A la derecha la rodaja con mayor activación cerebral.

Las áreas de color rojo muestran la activación de las regiones del cerebro involucradas en hacer la tarea motora propuesta a cada sujeto. Recordemos que dicha tarea consistía en pulsar con un dedo sobre un panel con botones de respuesta al ritmo de un tono auditivo. En las imágenes se ve que el hemisferio izquierdo muestra una activación mayor. Observamos que GP y SVM tienen mayor focalidad, lo que quiere decir que el área activada es más pequeña y está más focalizada en una región.

ESTUDIO 2

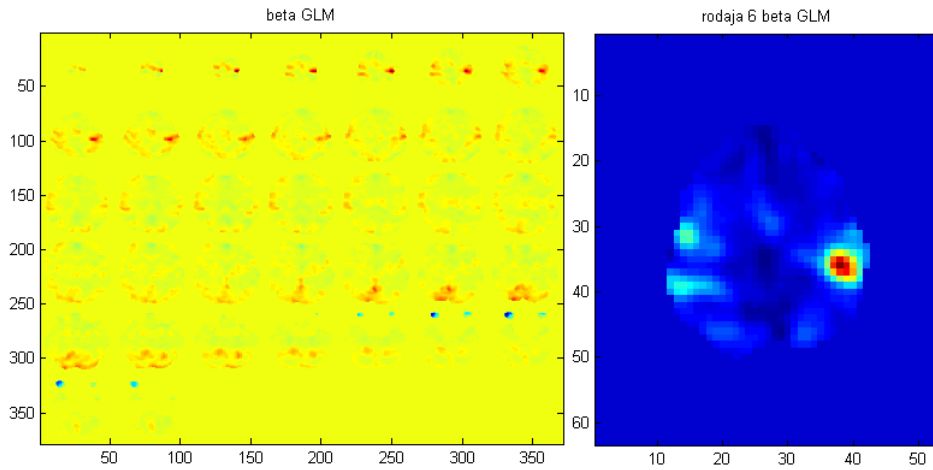


Figura 3.11: Mapa de activación de β usando **GLM** del estudio2. A la derecha la rodaja más significativa en la tarea motora.

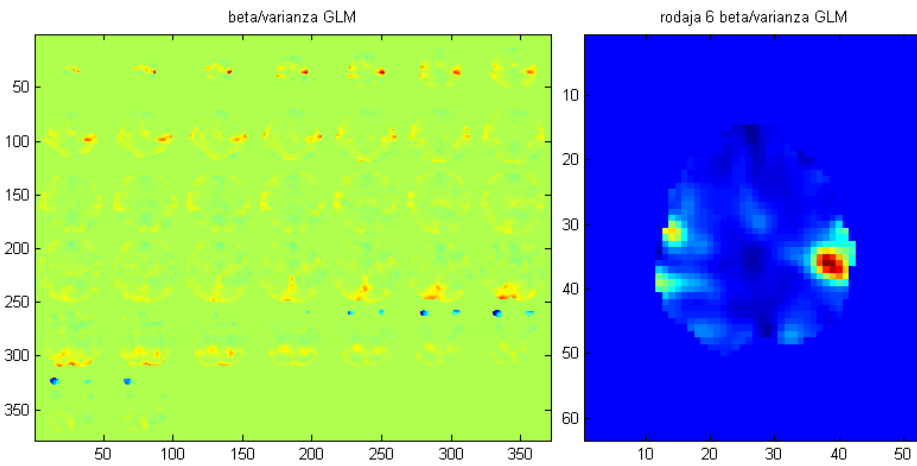


Figura 3.12: Mapa de activación del cociente entre el vector de pesos y la varianza usando **GLM** del estudio 2. A la derecha la rodaja con mayor activación cerebral.

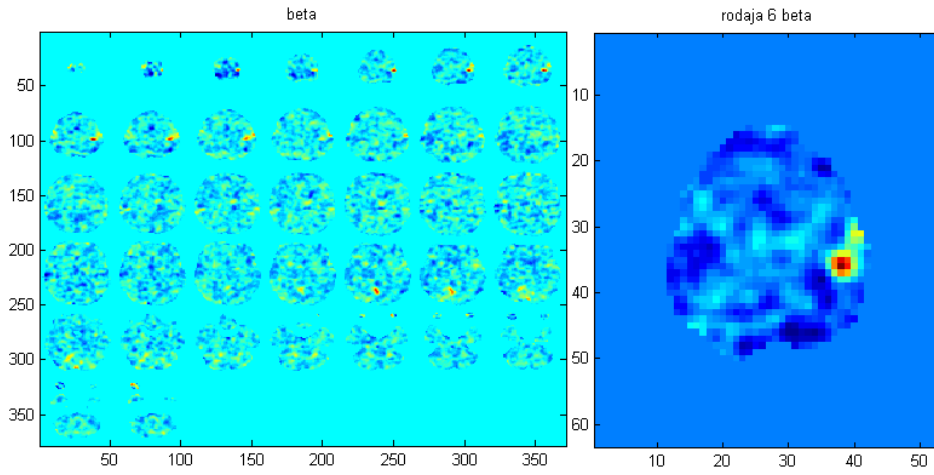


Figura 3.13: Mapa de activación de β usando GP del estudio 2. A la derecha la rodaja más significativa en la tarea motora.

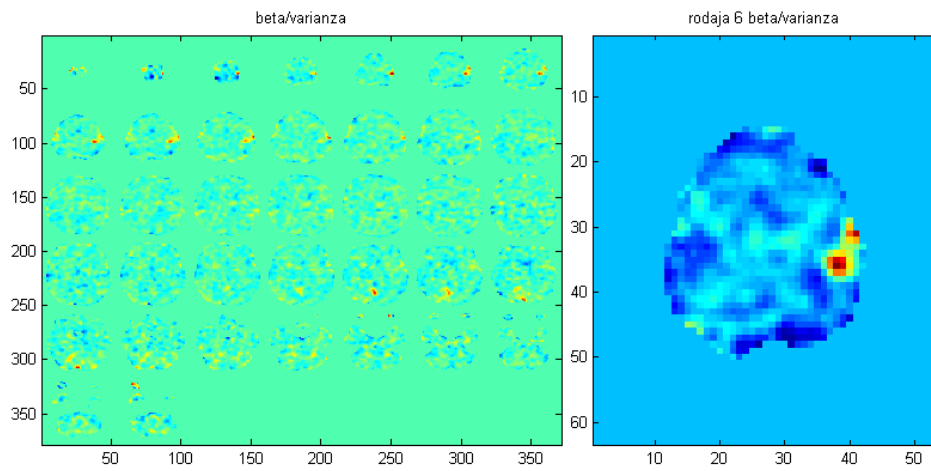


Figura 3.14: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 2. A la derecha la rodaja con mayor activación cerebral.

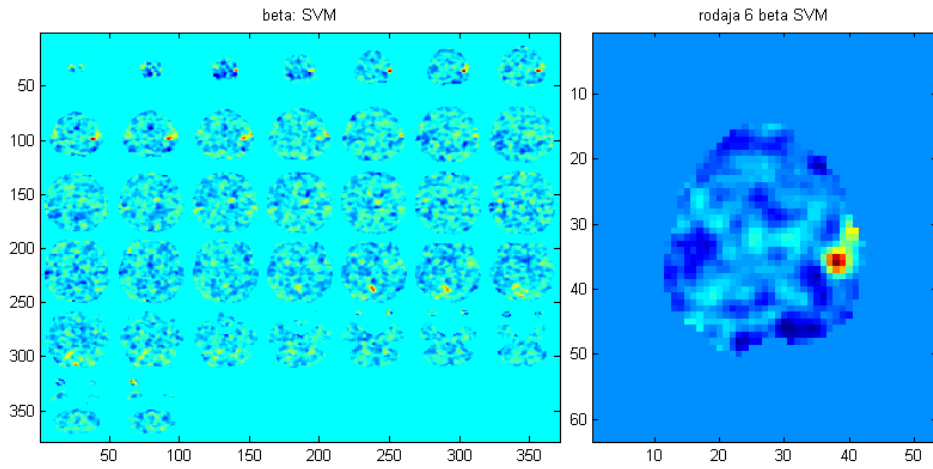


Figura 3.15: Mapa de activación de β usando SVM del estudio 2. A la derecha la rodaja más significativa en la tarea motora.

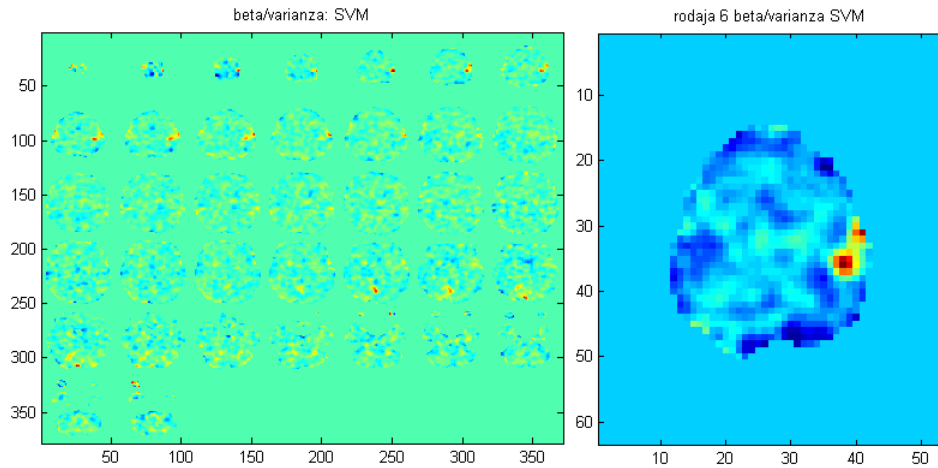


Figura 3.16: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 2. A la derecha la rodaja con mayor activación cerebral.

En el caso del estudio 2 se observa una mayor focalidad en GP y SVM, sin verse ningún tipo de artefacto.

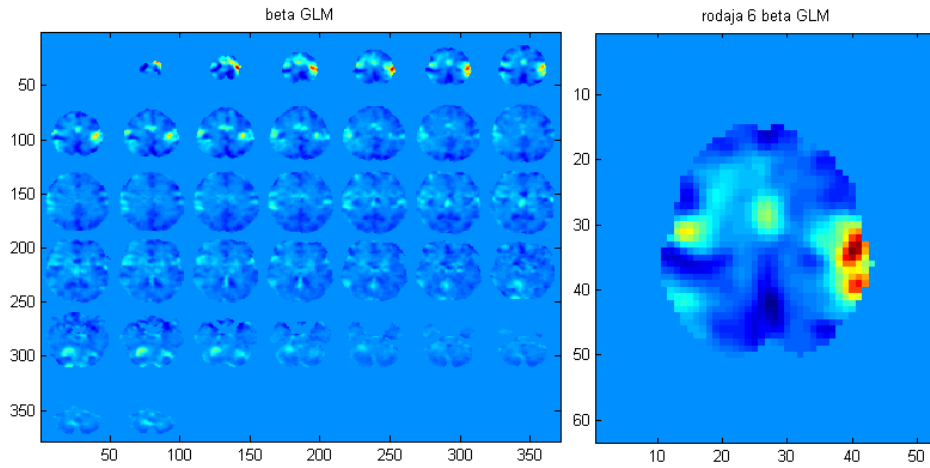
ESTUDIO 3

Figura 3.17: Mapa de activación de β usando *GLM* del estudio 3. A la derecha la rodaja más significativa en la tarea motora.

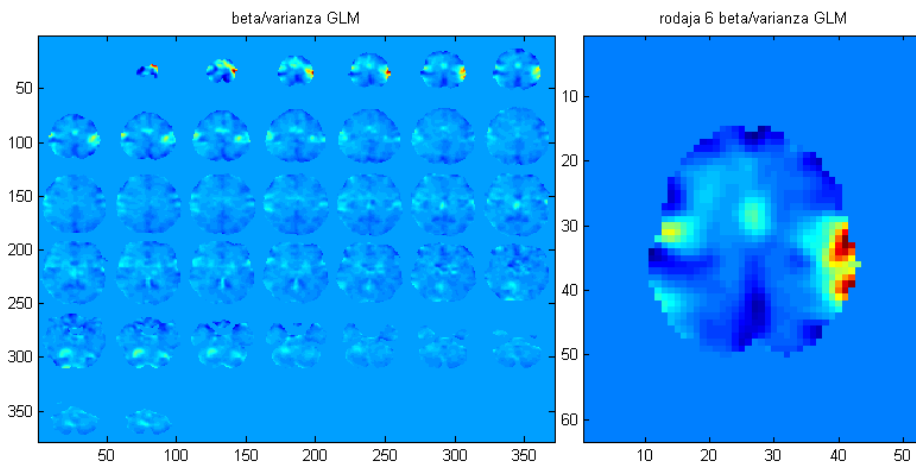


Figura 3.18: Mapa de activación del cociente entre el vector de pesos y la varianza usando *GLM* del estudio 3. A la derecha la rodaja con mayor activación cerebral.

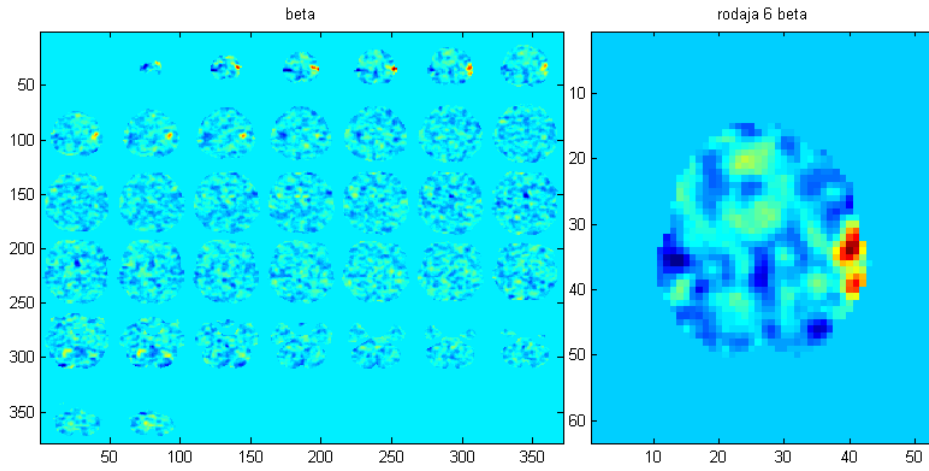


Figura 3.19: Mapa de activación de β usando GP del estudio 3. A la derecha la rodaja más significativa en la tarea motora.

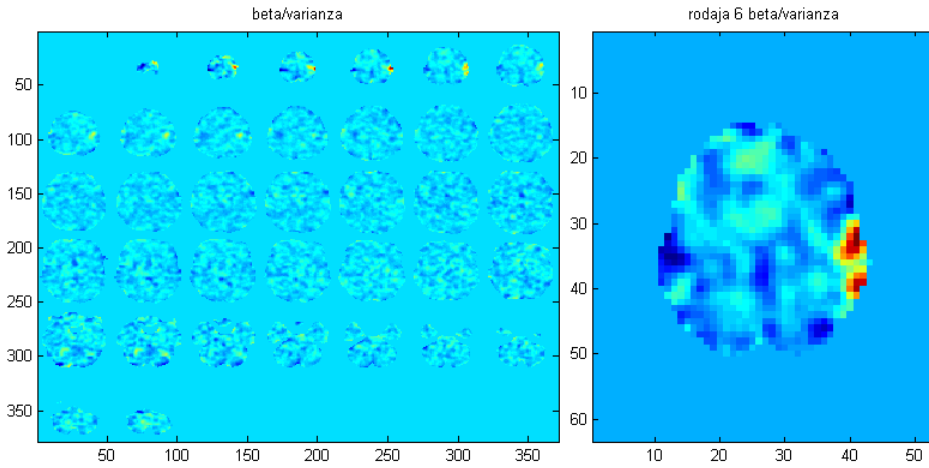


Figura 3.20: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 3. A la derecha la rodaja con mayor activación cerebral.

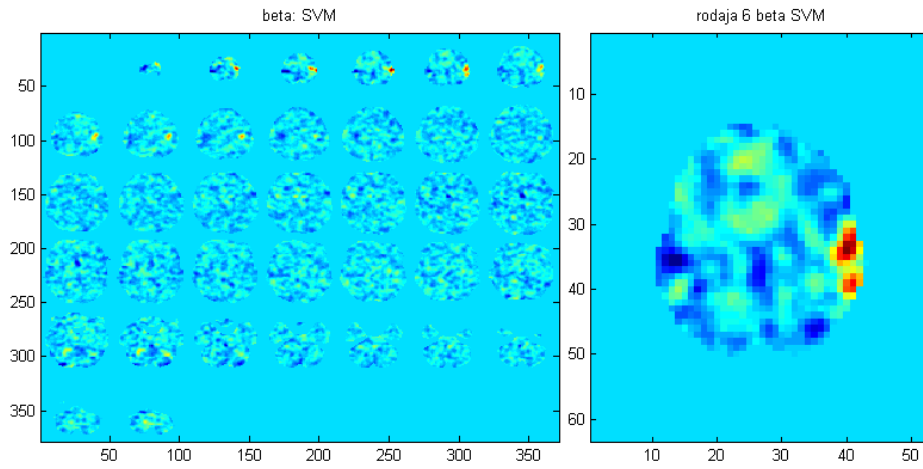


Figura 3.21: Mapa de activación de β usando SVM del estudio 3. A la derecha la rodaja más significativa en la tarea motora.

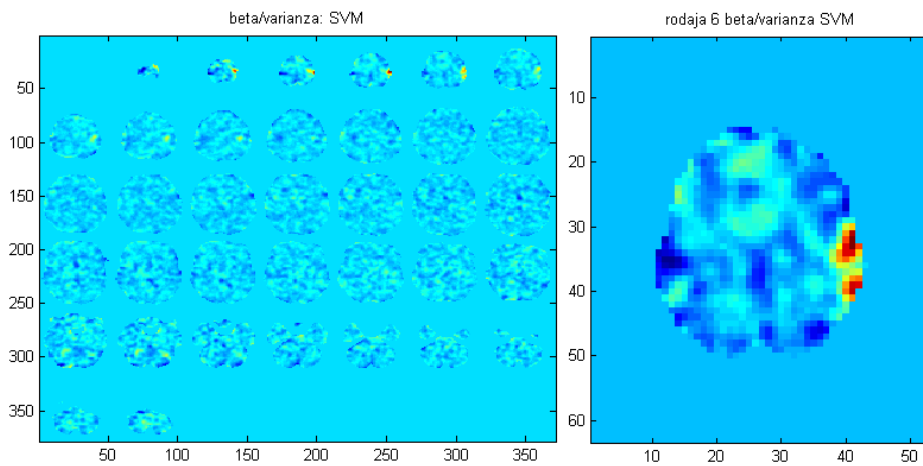


Figura 3.22: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 3. A la derecha la rodaja con mayor activación cerebral.

En el estudio 3 se ven los tres algoritmos muy parecidos, sin ningún tipo de artefacto. Vemos dos activaciones en el hemisferio izquierdo del cerebro.

ESTUDIO 4

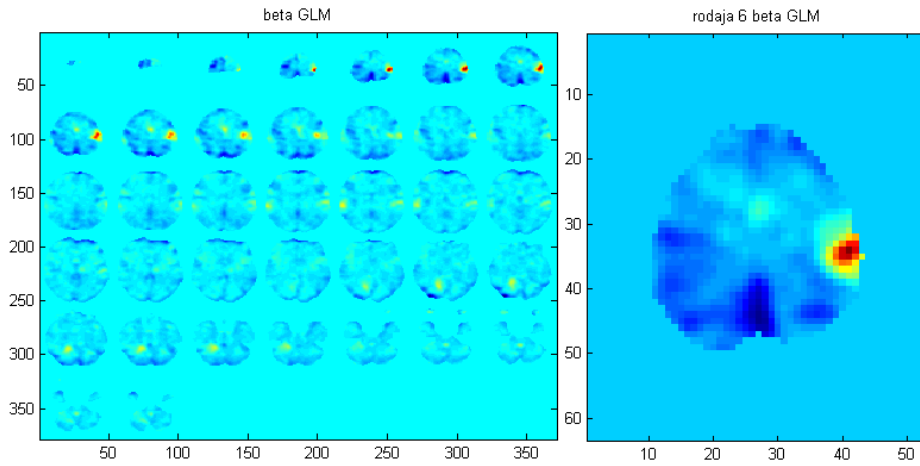


Figura 3.23: Mapa de activación de β usando *GLM* del estudio 4. A la derecha la rodaja más significativa en la tarea motora.

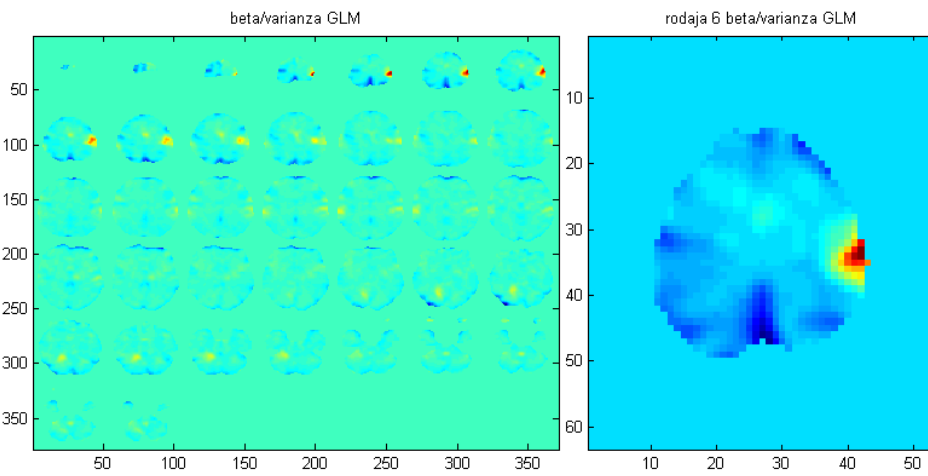


Figura 3.24: Mapa de activación del cociente entre el vector de pesos y la varianza usando *GLM* del estudio 4. A la derecha la rodaja con mayor activación cerebral.

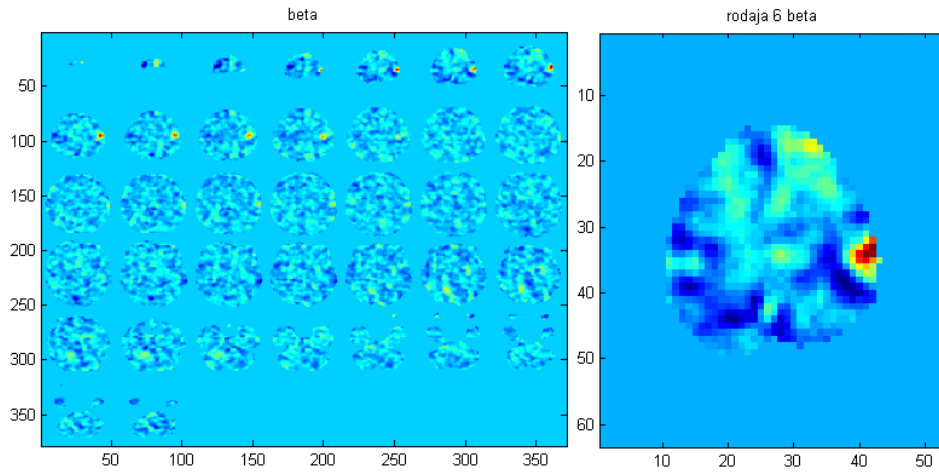


Figura 3.25: Mapa de activación de β usando GP del estudio 4. A la derecha la rodaja más significativa en la tarea motora.

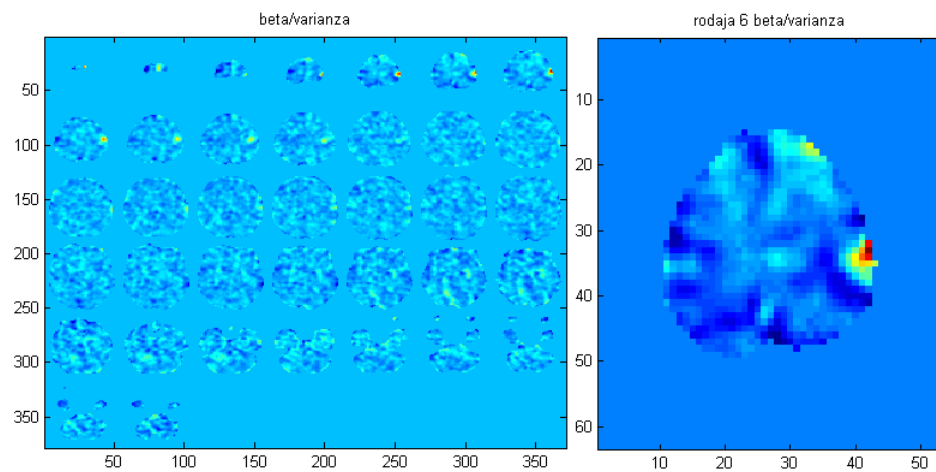


Figura 3.26: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 4. A la derecha la rodaja con mayor activación cerebral.

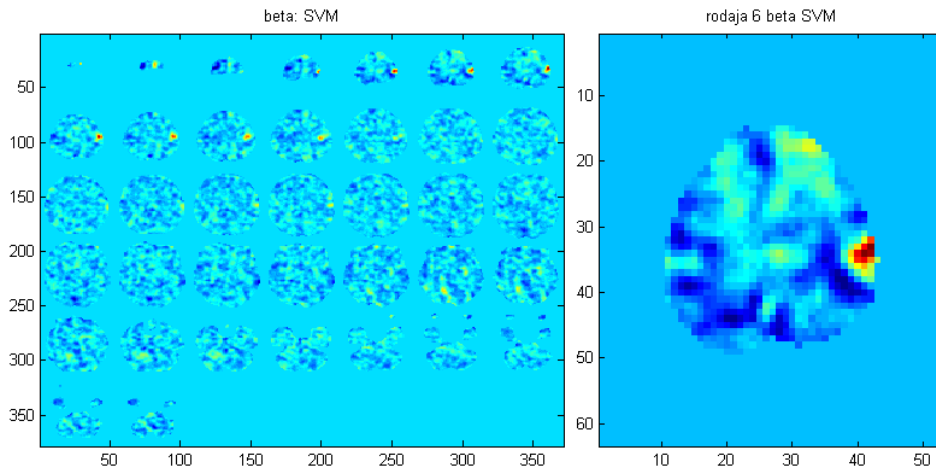


Figura 3.27: Mapa de activación de β usando SVM del estudio 4. A la derecha la rodaja más significativa en la tarea motora.

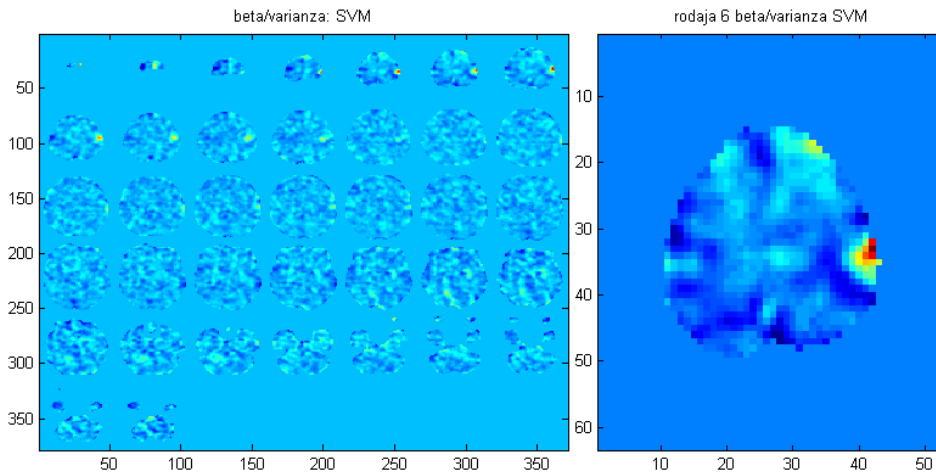


Figura 3.28: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 4. A la derecha la rodaja con mayor activación cerebral.

En el estudio 4 se observa mayor focalidad en los algoritmos multivariantes (GP y SVM) y también más ruido.

Las imágenes anteriores muestran el vector de pesos (β) y el cociente entre β y la varianza usando GLM, GP y SVM, todo ello en el caso de utilizar la matriz de covarianzas lineal. En los mapas del vector de pesos (3.5,3.11,3.17,3.23,3.7,3.13,3.19,3.25,3.9,3.15,3.21,3.27) y del cociente entre β y la varianza (3.6,3.12,3.18,3.24,3.8,3.14,3.20,3.26,3.10,3.16,3.22,3.28) se pueden ver áreas significantes de color rojo que indican que el vector de pesos β , es mayor. El ruido en los mapas de activación se limpia diviendo β entre la varianza. Lo que conseguimos con estos mapas (β /varianza) es enfatizar aquellos voxels que tienen mayor significancia. Para ver el área de activación más cerca, a la derecha de las imágenes observamos la rodaja 6 donde apreciamos una mayor activación cerebral en el lóbulo frontal del lado izquierdo del cerebro, debido a la tarea motora que los sujetos tenían que realizar. Recordemos que los mapas representados presentan 37 cortes o rodajas distintos del cerebro.

3.4.2. Caso 2: Matriz de covarianzas no lineal

A continuación se muestran los resultados obtenidos en cuatro estudios en el caso de utilizar la matriz de covarianzas no lineal. Hemos elegido los mismos cuatro estudios que en el caso lineal para observar las diferencias. Los resultados que se muestran hacen referencia a los algoritmos multivariantes GP y SVM.

ESTUDIO 1

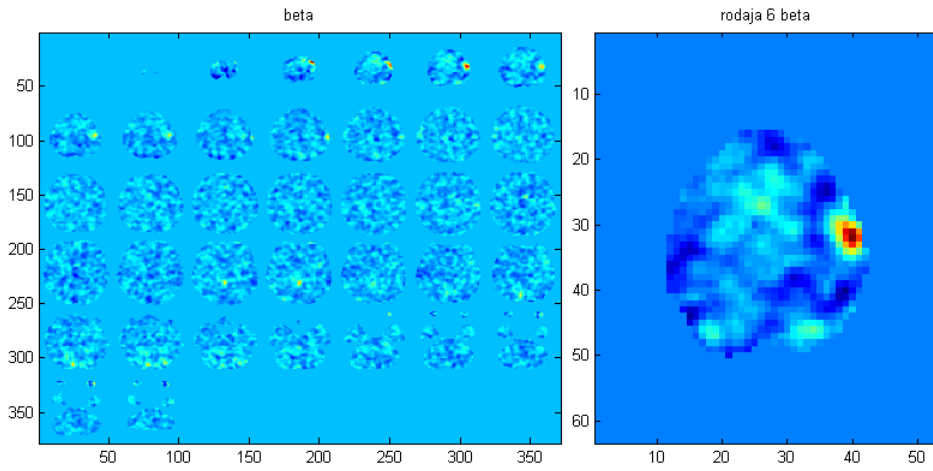


Figura 3.29: Mapa de activación de β usando GP del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

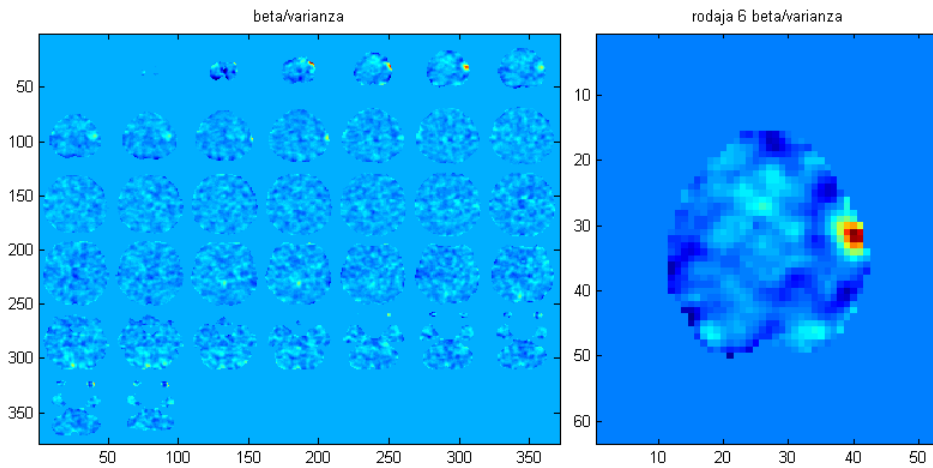


Figura 3.30: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 1. A la derecha la rodaja con mayor activación cerebral.

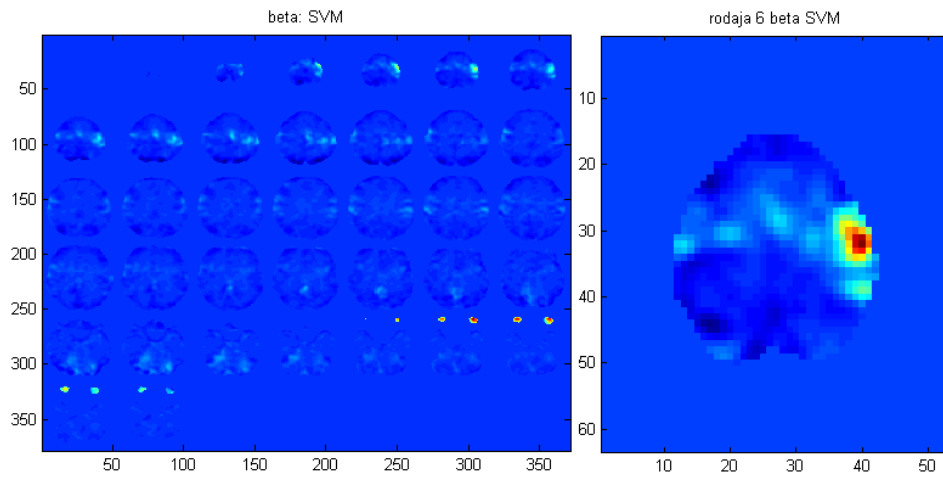


Figura 3.31: Mapa de activación de β usando SVM del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

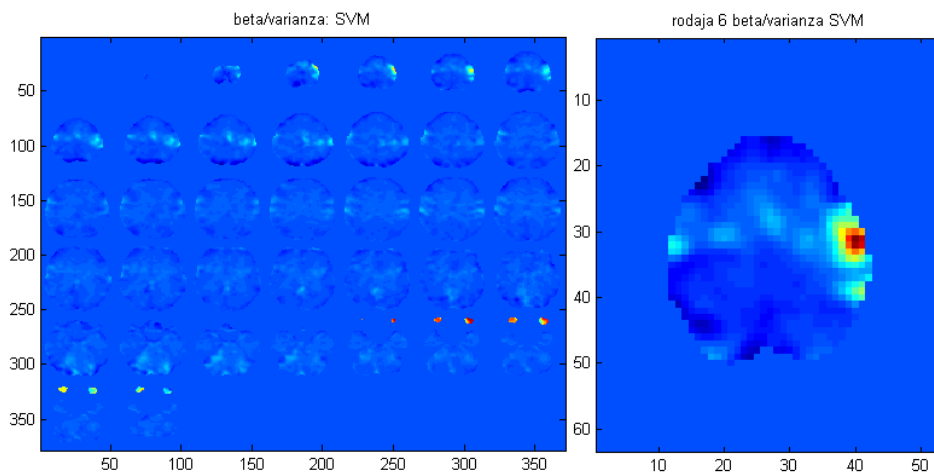


Figura 3.32: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 1. A la derecha la rodaja con mayor activación cerebral.

Las áreas de color rojo muestran la activación de las regiones del cerebro involucradas en hacer la tarea motora propuesta a cada sujeto. Recordemos que dicha tarea consistía en pulsar con un dedo sobre un panel con botones de respuesta. En las imágenes se ve que el hemisferio izquierdo muestra una activación mayor.

ESTUDIO 2

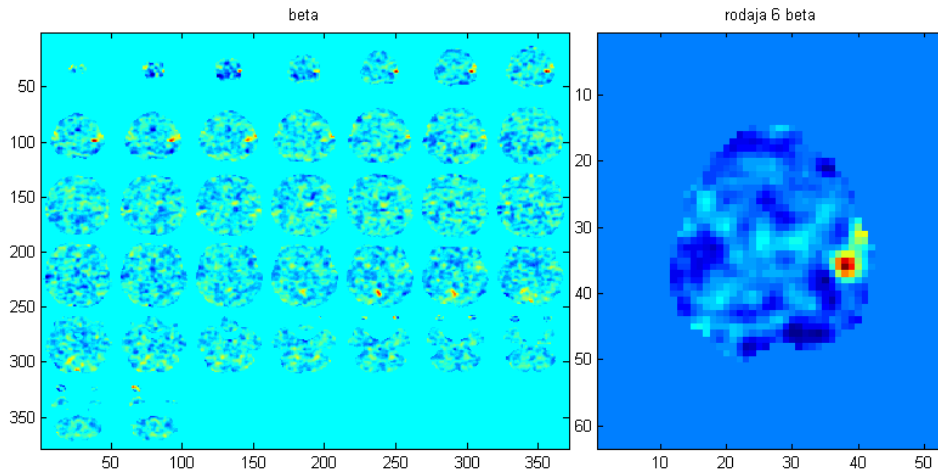


Figura 3.33: Mapa de activación de β usando **GP** del estudio 2. A la derecha la rodaja más significativa en la tarea motora.

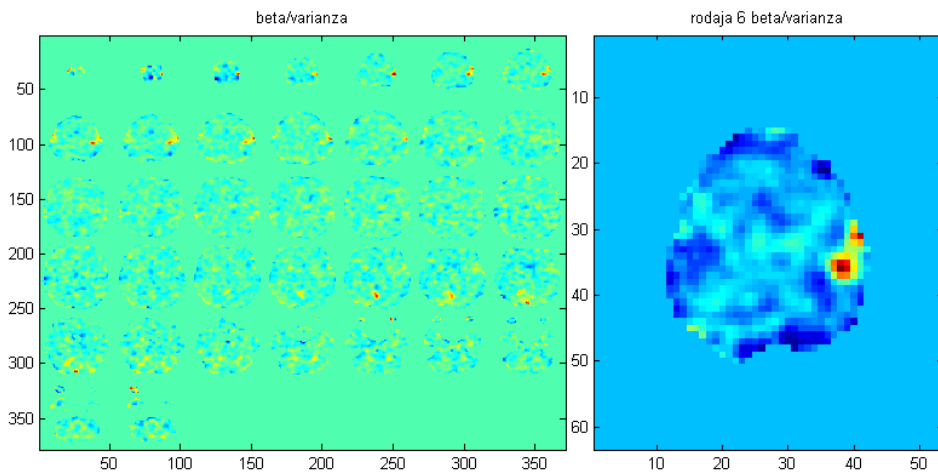


Figura 3.34: Mapa de activación del cociente entre el vector de pesos y la varianza usando **GP** del estudio 2. A la derecha la rodaja con mayor activación cerebral.

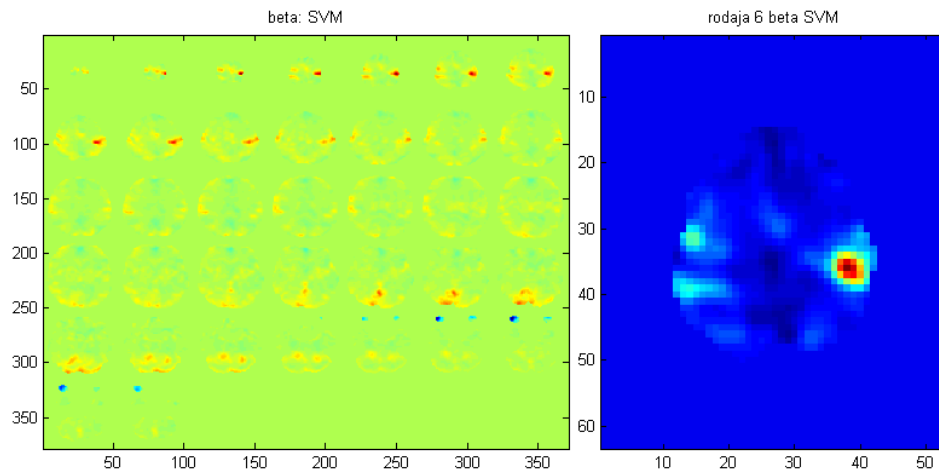


Figura 3.35: Mapa de activación de β usando SVM del estudio 2. A la derecha la rodaja más significativa en la tarea motora.

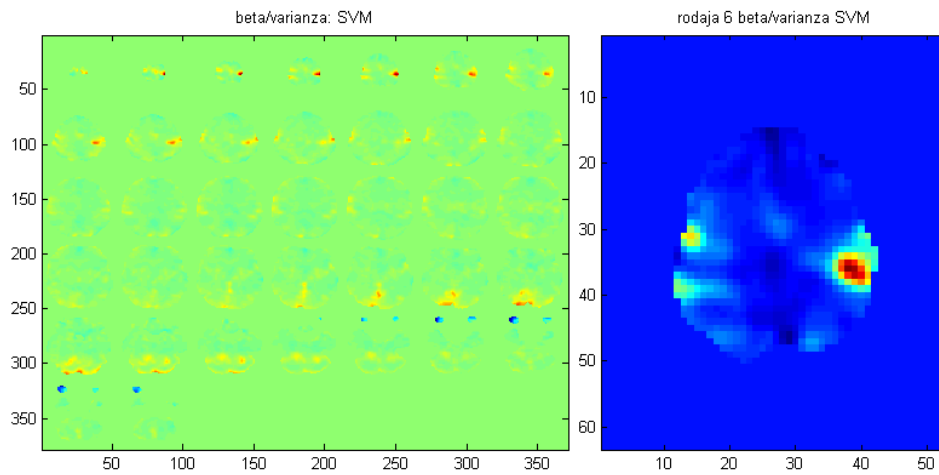


Figura 3.36: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 2. A la derecha la rodaja con mayor activación cerebral.

ESTUDIO 3

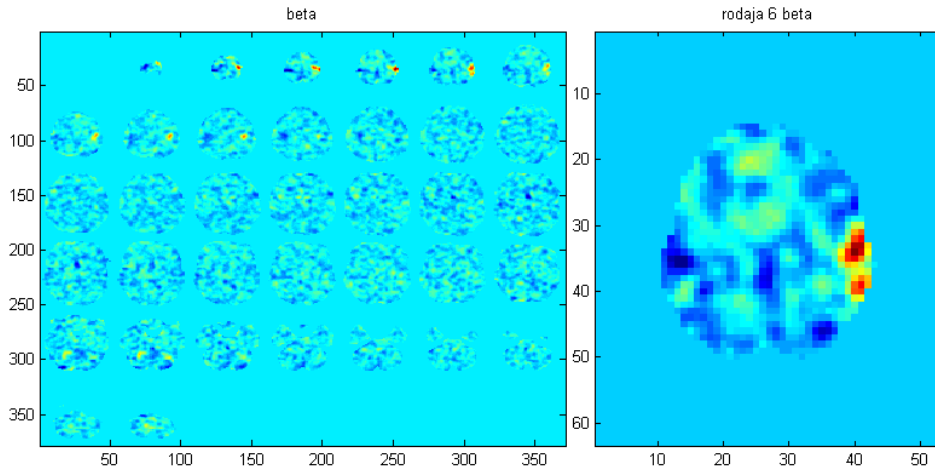


Figura 3.37: Mapa de activación de β usando GP del estudio 3. A la derecha la rodaja más significativa en la tarea motora.

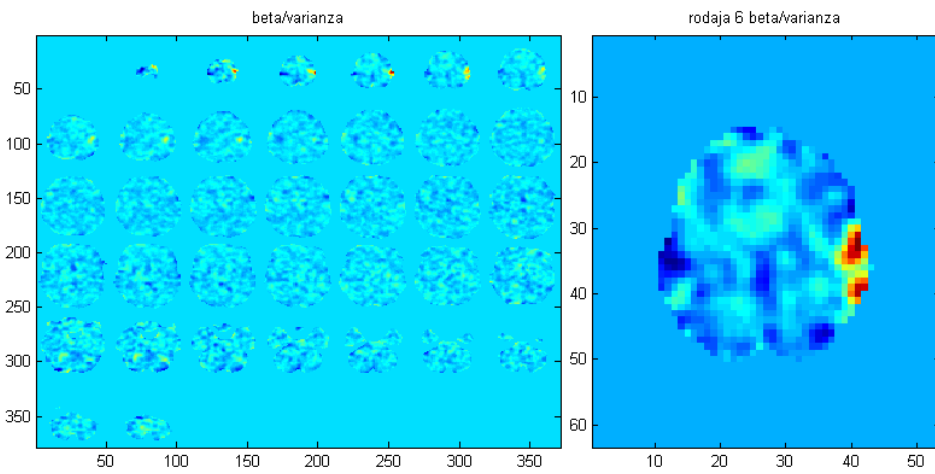


Figura 3.38: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 3. A la derecha la rodaja con mayor activación cerebral.

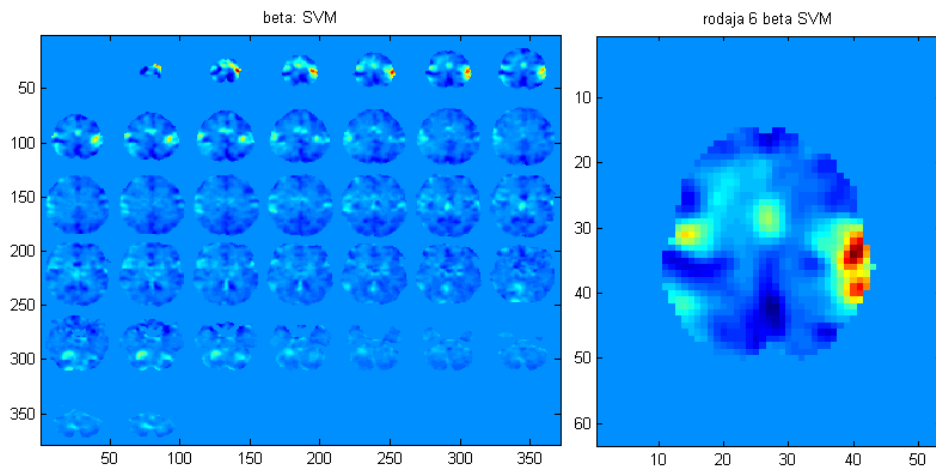


Figura 3.39: Mapa de activación de β usando SVM del estudio 3. A la derecha la rodaja más significativa en la tarea motora.

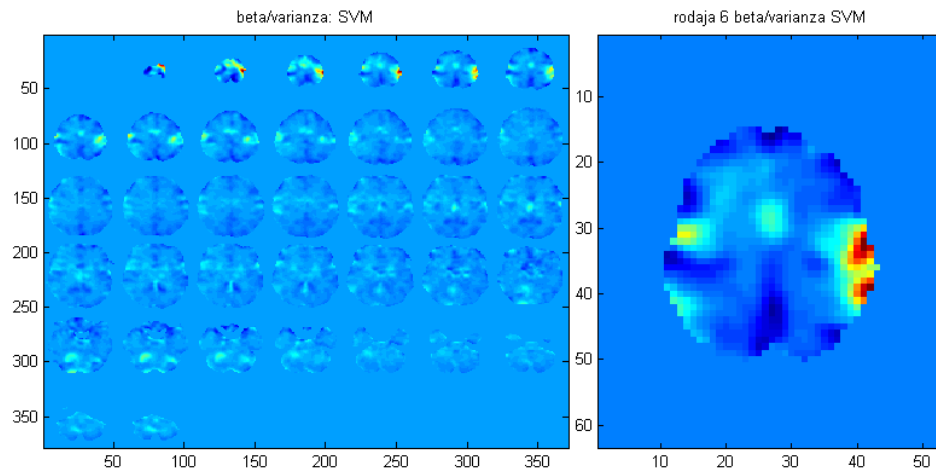


Figura 3.40: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 3. A la derecha la rodaja con mayor activación cerebral.

ESTUDIO 4

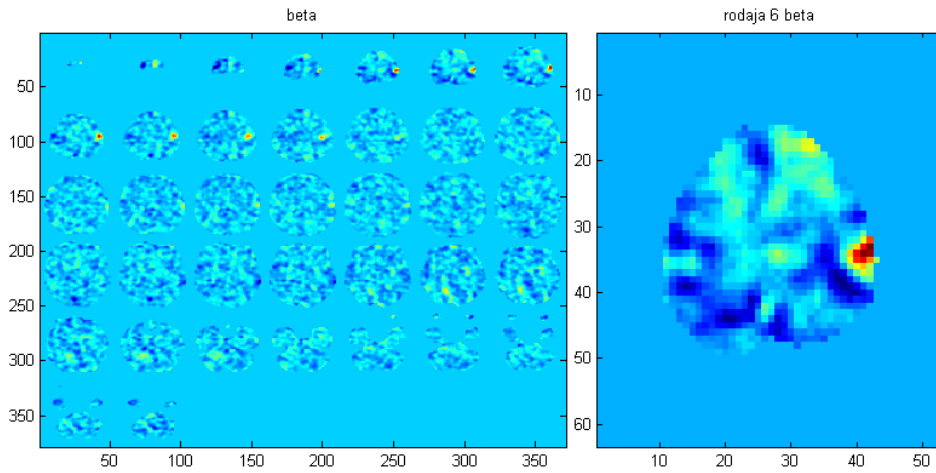


Figura 3.41: Mapa de activación de β usando GP del estudio 4. A la derecha la rodaja más significativa en la tarea motora.

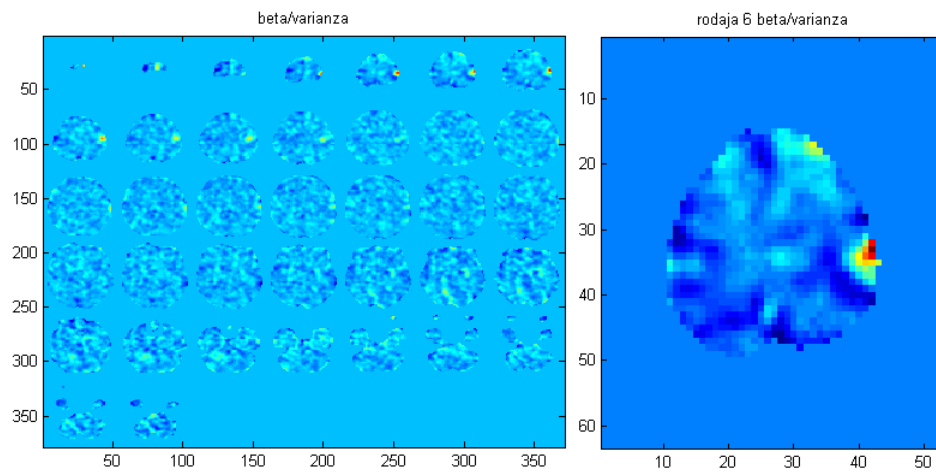


Figura 3.42: Mapa de activación del cociente entre el vector de pesos y la varianza usando GP del estudio 4. A la derecha la rodaja con mayor activación cerebral.

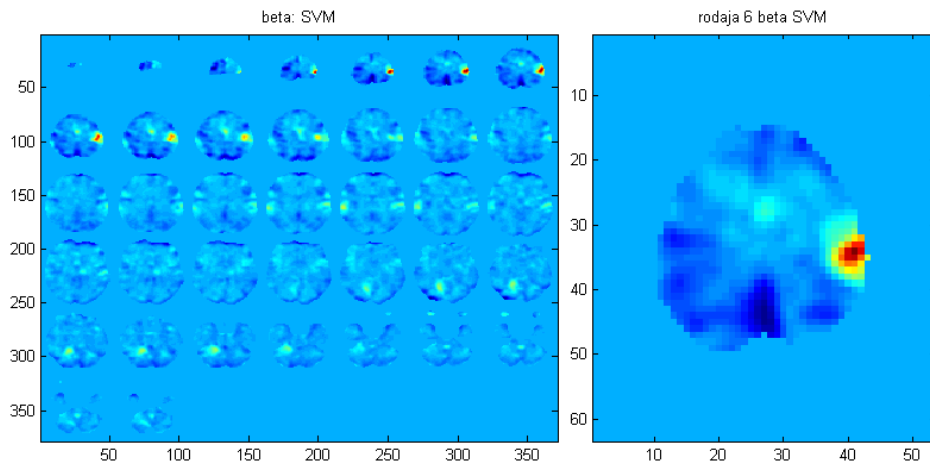


Figura 3.43: Mapa de activación de β usando SVM del estudio 4. A la derecha la rodaja más significativa en la tarea motora.

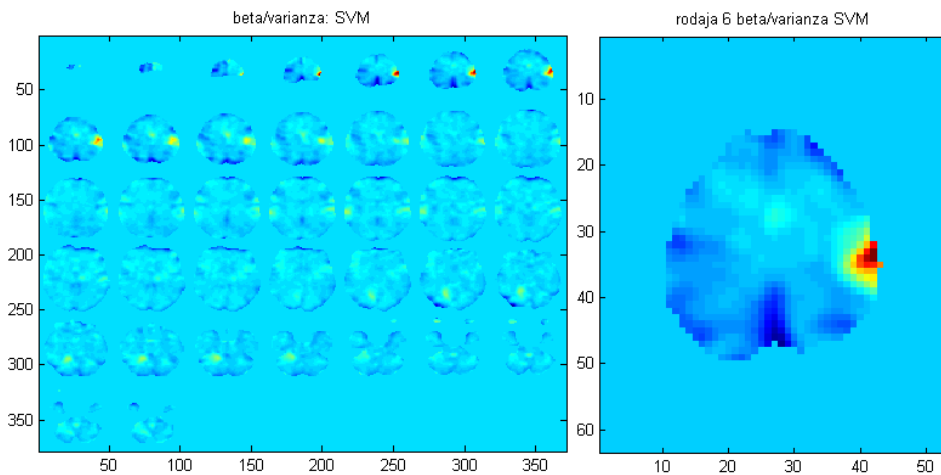


Figura 3.44: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM del estudio 4. A la derecha la rodaja con mayor activación cerebral.

~~08~~ CAPÍTULO 3. DESCRIPCIÓN DE LOS ALGORITMOS, DATOS Y RESULTADOS

Las imágenes anteriores muestran el vector de pesos, (β) y el cociente entre β y la varianza usando GLM, GP y SVM, todo ello en el caso de utilizar la matriz de covarianzas no lineal (caso 2). En los mapas de activación del vector de pesos (3.29, 3.33, 3.37, 3.41, 3.31, 3.35, 3.39, 3.43) y del cociente entre β y la varianza (3.30, 3.34, 3.38, 3.42, 3.32, 3.36, 3.40, 3.44) se pueden ver áreas significantes de color rojo que indican que el vector de pesos β , es mayor. Para ver dicha área más cerca, a la derecha de estas imágenes observamos una de las rodajas donde apreciamos una mayor activación cerebral, debido a la tarea motora que los sujetos realizaban.

Los resultados obtenidos en el caso de utilizar la matriz de covarianzas lineal son similares a los del caso de usar la matriz de covarianzas no lineal, por tanto, en el caso lineal no se observa mejora con respecto al caso no lineal. En ambos casos vemos una activación en el lóbulo frontal inmediatamente anterior al surco central del hemisferio izquierdo del cerebro debido a la tarea motora realizada por los sujetos. Los estudios son bastante consistentes en los dos casos.

A continuación representamos los resultados obtenidos variando el parámetro “ ϵ ” de las SVM. Para ello usamos los valores 0.1, 0.3, 0.5 y 0.6 y vemos las diferencias existentes una vez vamos aumentando el valor de ϵ , tanto en el caso de utilizar la matriz de covarianzas lineal como no lineal.

Comenzamos viendo los resultados en el caso de emplear la **matriz de covarianzas lineal**.

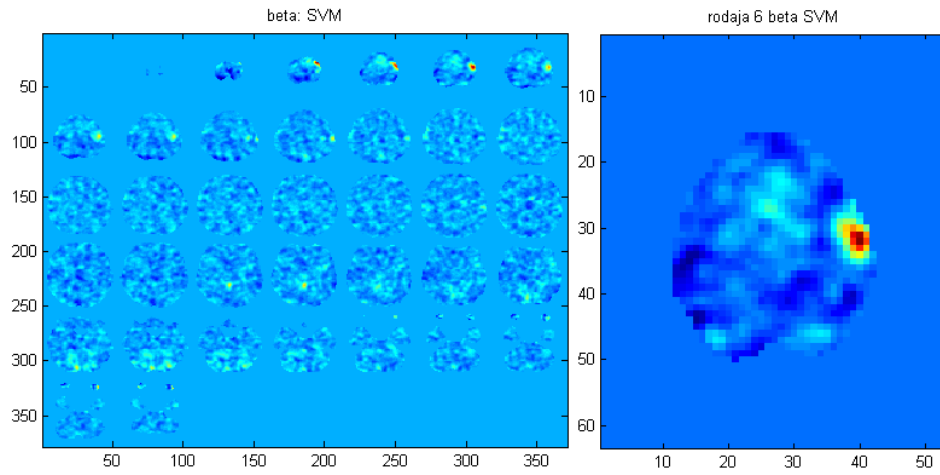
EPSILON 0.1

Figura 3.45: Mapa de activación de β usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

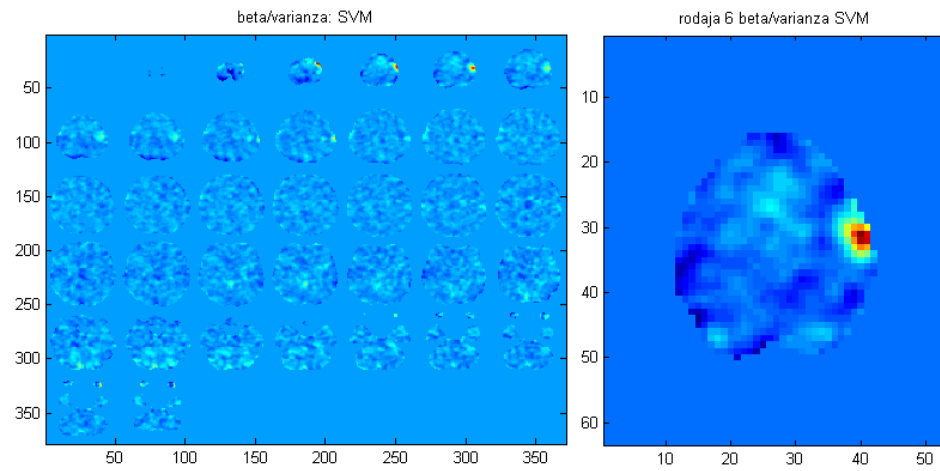


Figura 3.46: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

EPSILON 0.3

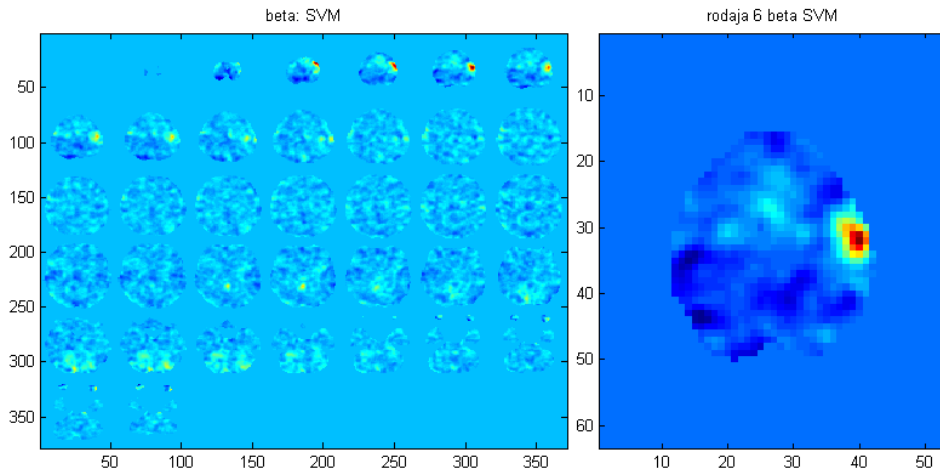


Figura 3.47: Mapa de activación de β usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

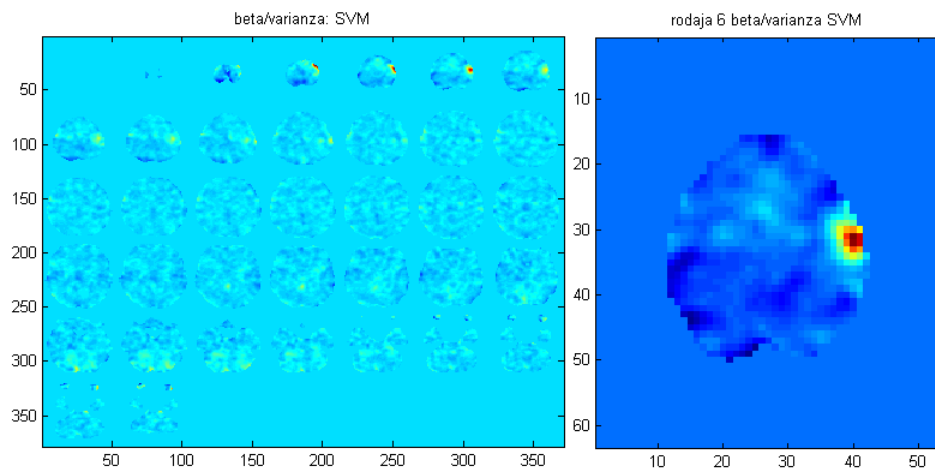


Figura 3.48: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

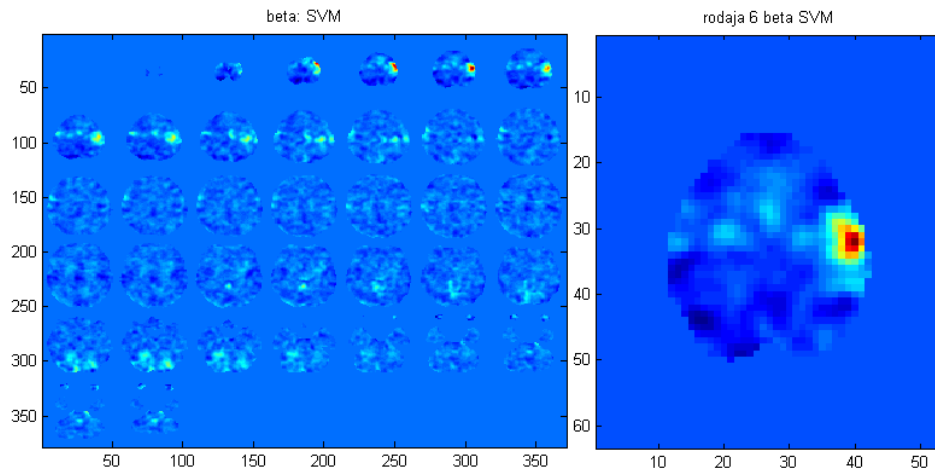
EPSILON 0.5

Figura 3.49: Mapa de activación de β usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

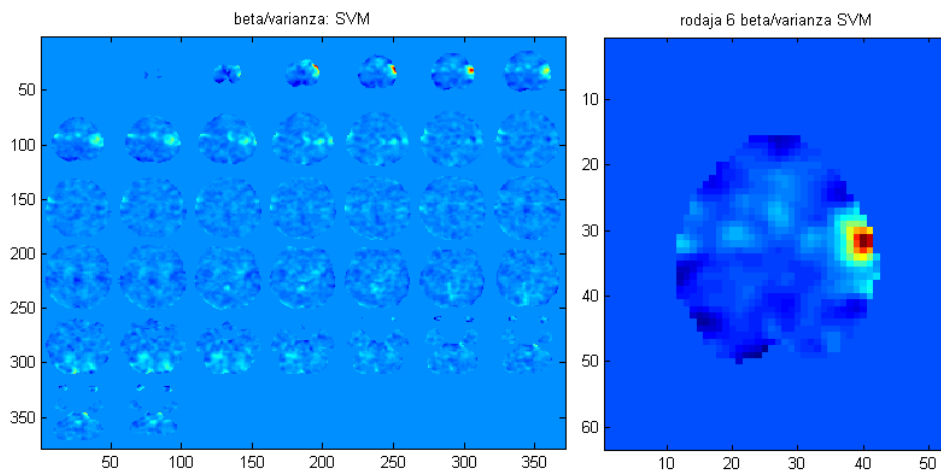


Figura 3.50: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

EPSILON 0.6

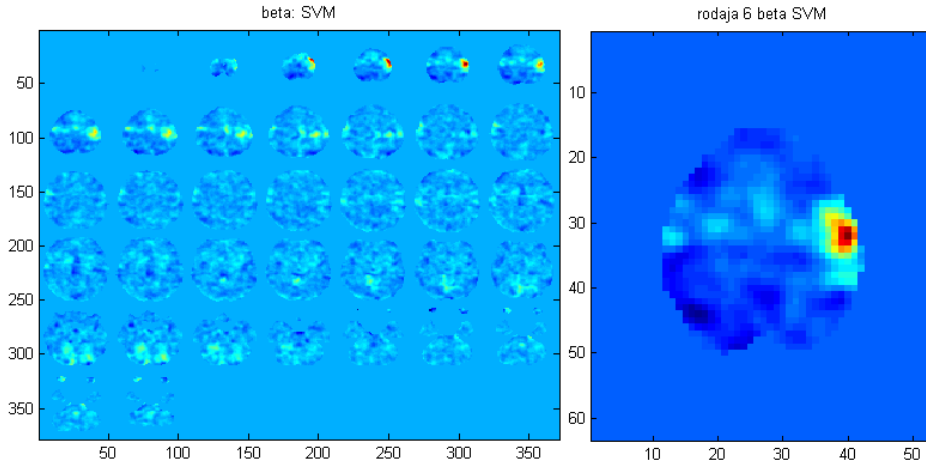


Figura 3.51: Mapa de activación de β usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

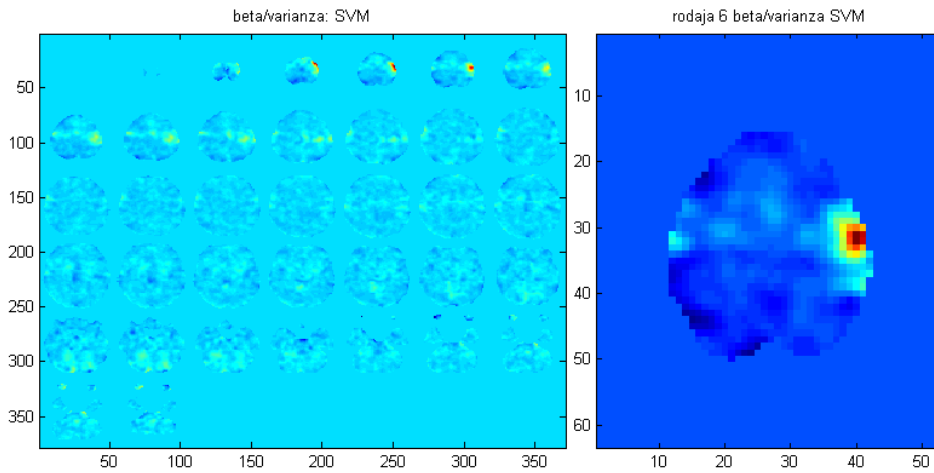


Figura 3.52: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

A mayor valor de ϵ , utilizamos menos vectores de pesos y por tanto menos información. Por consiguiente se observa las imágenes más suavizadas y con menos artefactos.

A continuación vemos los resultados en el caso de emplear la **matriz de covarianzas no lineal**.

EPSILON 0.1

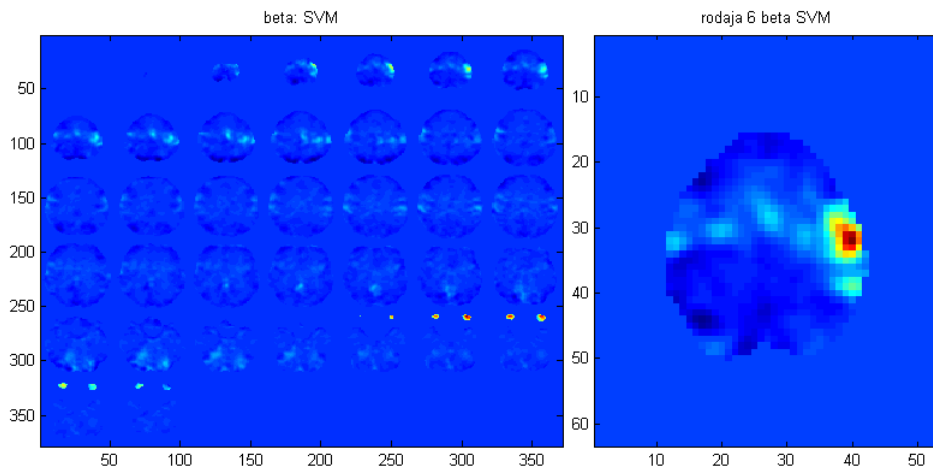


Figura 3.53: Mapa de activación de β usando **SVM** con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

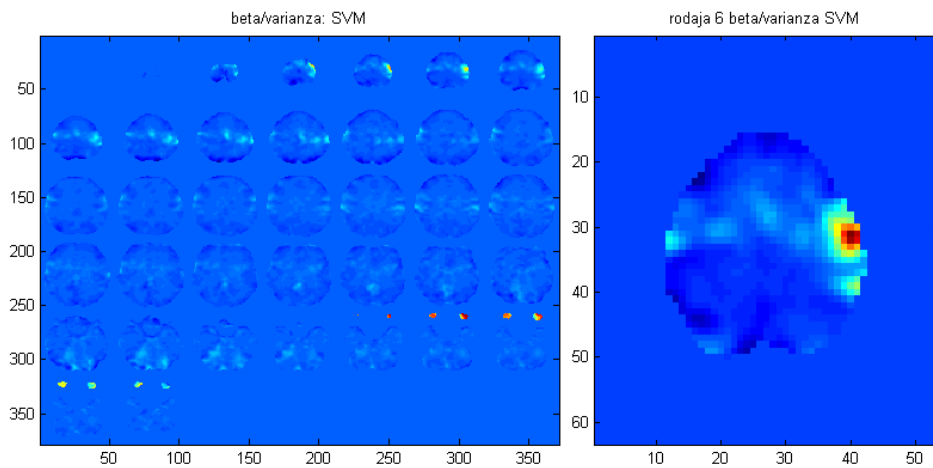


Figura 3.54: Mapa de activación del cociente entre el vector de pesos y la varianza usando **SVM** con valor de $\epsilon=0.1$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

EPSILON 0.3

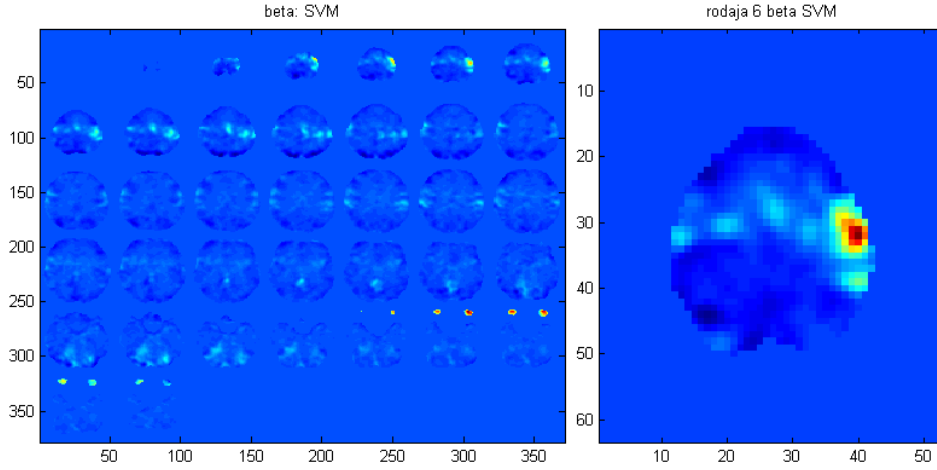


Figura 3.55: Mapa de activación de β usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

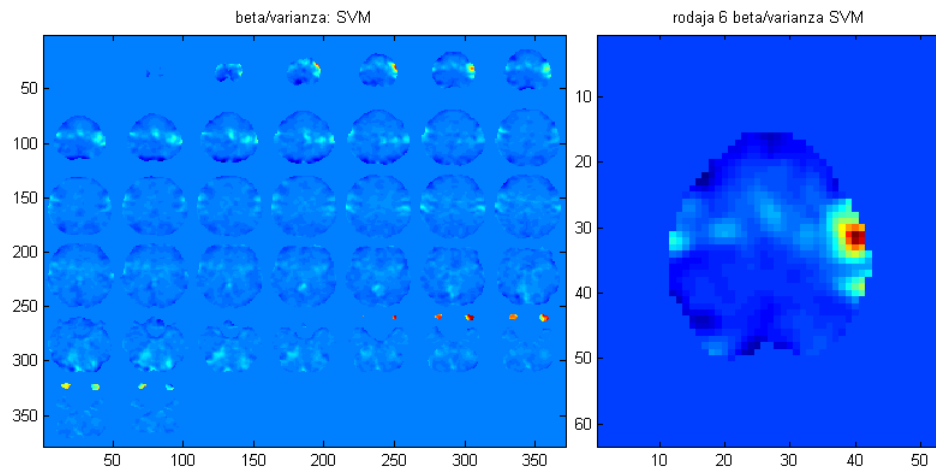


Figura 3.56: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.3$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

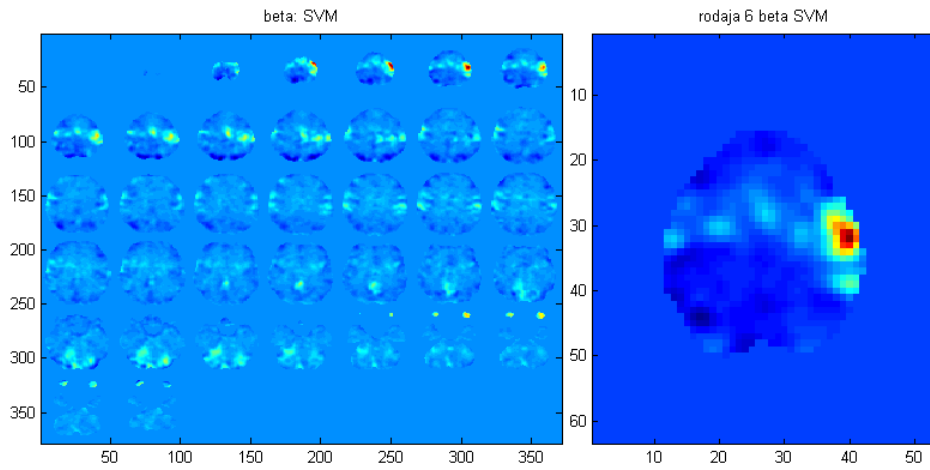
EPSILON 0.5

Figura 3.57: Mapa de activación de β usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

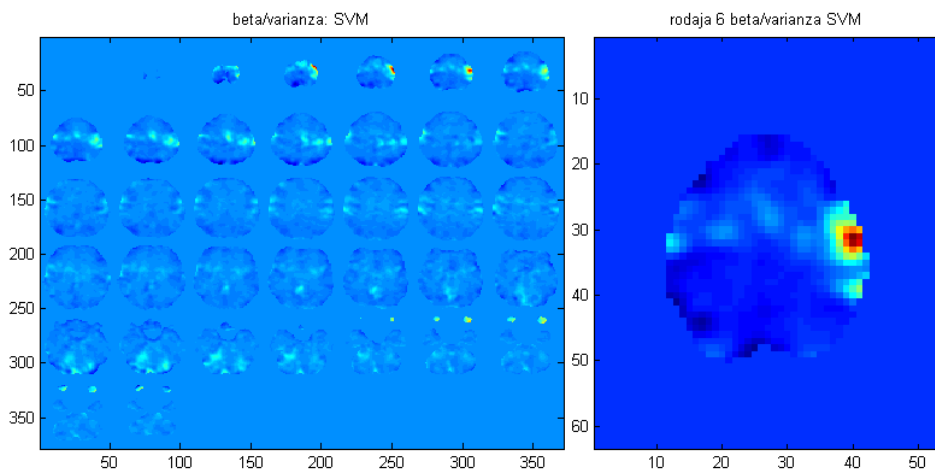


Figura 3.58: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.5$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

EPSILON 0.6

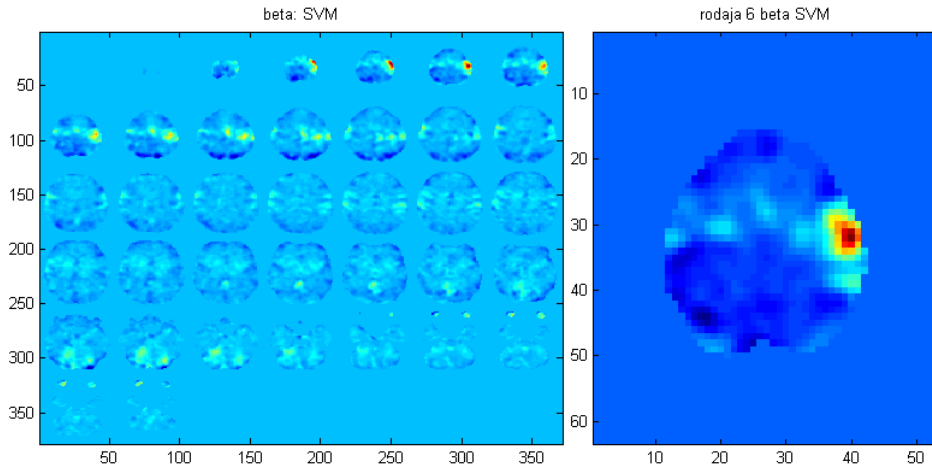


Figura 3.59: Mapa de activación de β usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja más significativa en la tarea motora.

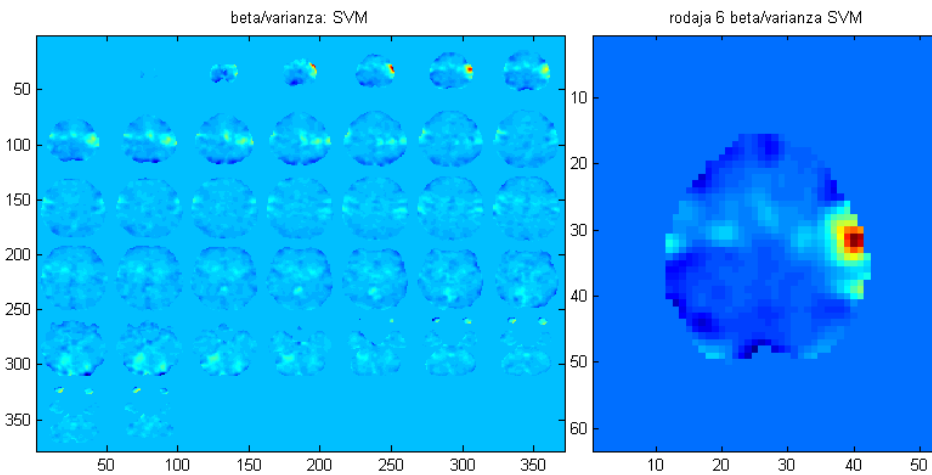


Figura 3.60: Mapa de activación del cociente entre el vector de pesos y la varianza usando SVM con valor de $\epsilon=0.6$ del estudio 1. A la derecha la rodaja con mayor activación cerebral.

En el caso no lineal, al igual que el caso lineal se observa que a menor valor de ϵ , la imagen se ve más ruidosa debido a que utilizamos más vectores de peso.

CONCLUSIONES

4.1. Discusión

Se han presentado un conjunto de algoritmos tanto univariantes como multivariantes. Dentro de los univariantes hemos visto el método lineal general (GLM) y dentro de los multivariantes las Redes de Procesos Gaussianos (GPR) y las Máquinas de Vectores Soporte (SVM). Debido a la alta dimensionalidad de los datos dentro de las imágenes de Resonancia Magnética Funcional utilizamos el método dual para su simplificación.

Dichos algoritmos se han probado con datos de entrada basados en Imágenes de Resonancia Magnética Funcional, obteniéndose una serie temporal de imágenes tridimensionales que indican la variación de la oxigenación en el cerebro. Para ello fueron estudiados diez sujetos sanos (todos diestros) usando un escáner Siemens de 3.0 Teslas. Los estímulos fueron presentados mediante unos auriculares compatibles con la máquina de Resonancia Magnética. El paradigma consistía en un bloque de 20 segundos de actividad y 20 segundos de reposo. A los sujetos se les pidió que tocasen con un dedo un panel con botones de respuesta al ritmo de un tono auditivo.

Se observa que los métodos multivariantes producen mejores mapas β que los univariantes, ya que en los primeros para cada valor de β_i se usan todos los voxels mientras que en los segundos cada β_i depende únicamente del voxel i . Además en el caso de GP y SVM se ve mayor focalidad en las imágenes.

También podemos observar las diferencias que existen variando el parámetro ϵ de las SVM. Se sabe que a mayor ϵ , menor cantidad de datos utilizamos para obtener el mapa de activación, por tanto se obtendrán mapas β más suavizados y menos ruidosos.

También hemos visto que en el caso lineal no se observa mejora con respecto al caso no lineal.

Se ha presentado un método de cálculo de la varianza de los voxels, con moderado coste computacional justificado teóricamente.

4.2. Trabajos futuros

Algunos ejemplos de caminos futuros y aplicaciones de la tecnología de resonancia magnética funcional son:

Extensión del t-map al caso no lineal.

Introducción de este software en el paquete de aplicaciones MRI SPM [[26]], que es un software de distribución gratuita que inferencia estadísticamente datos de Resonancia Magnética.

Se podría probar lo que se ha realizado en este proyecto mediante otros tipos de kernels. El kernel que se ha utilizado en este caso ha sido de base radial gaussiana. La investigación en kernels de Mercer está abierta y produce gran cantidad de comunicaciones científicas al año. No se puede, por tanto, descartar el uso de kernels que produzcan mejores prestaciones que los aquí utilizados.

Aplicaciones a otros campos relacionados, tales como electroencefalografía y magnetoencefalografía (EEG y MEG).

PRESUPUESTO DEL PROYECTO

En este apéndice se presentan justificados los costes globales de la realización de este Proyecto Fin de Carrera. Tales costes, imputables a gastos de personal y de material, se pueden deducir de las Tablas A.1, A.2 y A.4.

En la Tabla A.1 se muestran las fases del proyecto y el tiempo dedicado para cada una de ellas. Así pues, el tiempo total dedicado por el proyectando ha sido de 1.040 horas, de las cuales aproximadamente un 20 % han sido compartidas con el tutor del proyecto, por lo que el total asciende a 1.248 horas. Teniendo en cuenta que la tabla de honorarios del Colegio Oficial de Ingenieros Técnicos de Telecomunicación establece unas tarifas de 60 €/hora (teniendo en cuenta que son horas dentro de la jornada laboral, tal y como viene en el punto 19 de la baremación hecha por el COIT) el coste de personal, teniendo en cuenta sólo el trabajo del ingeniero de Telecomunicación, se sitúa en 93.600 €. A esto hay que añadir en el apartado de costes de personal los investigadores senior, el médico y los técnicos que han de estar presentes en cada experimento (en este proyecto se ha trabajado con 160 experimentos). El coste de personal aparece desglosado en la tabla A.2.

En la Tabla A.4 se recogen los costes de material desglosados en equipo informático, escáner, local de trabajo, documentación y gastos varios no atribuibles (material fungible, llamadas telefónicas, desplazamientos...). Ascienden, pues, a un total de 10.700 €.

A partir de estos datos, el presupuesto total es el mostrado en la Tabla A.5.

Tabla A.1: Fases del Proyecto

Fase 1	<i>Documentación</i>	280 horas
Fase 2	<i>Realización de las pruebas</i>	110 horas
Fase 3	<i>Implementación de los métodos propuestos</i>	400 horas
Fase 4	<i>Redacción de la memoria del proyecto</i>	250 horas
Fase 5	<i>Tiempo compartido con el tutor del proyecto</i>	208 horas
	TOTAL	1.248 horas

Tabla A.2: Costes de personal

Personal	Horas de trabajo	€/hora	Total (€)
<i>1 Ing.Telecomunicación</i>	1.248	60	74.880
<i>2 Investigadores</i>	2 horas/experimento	75 €/hora/pers	48.000
<i>1 Técnico</i>	2 horas/experimento	45 €/hora/pers	14.400
			137.280

Tabla A.3: Costes de material fungible

Equipo	Tiempo de uso	Precio	Total (€)
<i>Material para experimentos</i>		90 €	90
<i>Documentación</i>		250 €	250
			340

Tabla A.4: Costes de material inventariable

Equipo	Tiempo de uso	Precio	Total (€)
<i>Escáner</i>	5 horas (30 minutos/paciente)	600 €/hora	3.000
<i>Local</i>	12 meses	130 €/mes	1.560
<i>Ordenador gama alta</i>		2000 €	2.000
			6.560

Tabla A.5: Presupuesto

Concepto	Importe
Costes personal	137.280€
Costes material fungible	340 €
Costes material inventariable	6.560 €
Gastos indirectos	1.035€
TOTAL	145.215 €

MÁQUINAS DE VECTORES SOPORTE NO LINEALES

Para muchos problemas de clasificación o regresión, una solución lineal no da buenos resultados, en esos casos es necesaria una aproximación no lineal. Para poder generalizar los métodos empleados anteriormente a este caso en el que la función de decisión no es una función lineal de los datos, hay que tener en cuenta:

- Los datos aparecen en el problema de entrenamiento en forma de producto escalar: $x_i \cdot x_j$.
- Los datos que inicialmente pertenecen al espacio \mathbb{R}^d se mapean a otro espacio (posiblemente de dimensión infinita). Este espacio es el espacio euclídeo (\mathcal{H}), mediante ϕ [31] [32]. La aplicación es de la forma:

$$\phi : \mathbb{R}^d \mapsto \mathcal{H}$$

El algoritmo de entrenamiento depende de los datos a través de los productos en \mathcal{H} . Funciones de la forma: $\phi(x_i) \cdot \phi(x_j)$. Si hubiera un kernel K ([29], [30]), tal que $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, sólo se necesitaría usar K en el algoritmo de entrenamiento sin necesidad de saber qué es ϕ .

Ejemplo: $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{2\sigma^2}}$

En este ejemplo \mathcal{H} es de dimensión infinita por lo que no sería fácil trabajar con ϕ explícitamente. Sin embargo, si se reemplaza $x_i \cdot x_j$ por $K(x_i, x_j)$ en todas partes del algoritmo de entrenamiento, el algoritmo producirá una máquina de vectores soporte en un espacio infinito.

En la fase de test para poder usar esa máquina no hace falta calcular w (para ello habría que saber $\phi(s_i)$ explícitamente), basta con calcular el signo de:

$$f(x) = \sum_{i=1}^{N_s} \alpha_i y_i \phi(s_i) \cdot \phi(x) + b = \sum_{i=1}^{N_s} \alpha_i y_i K(s_i, x) + b \quad (\text{B.1})$$

donde s_i son los vectores soporte. Se evita calcular $\phi(x)$ explícitamente usando $K(s_i, x) = \phi(s_i) \cdot \phi(x)$.

Es fácil encontrar kernels tales que el algoritmo de entrenamiento y la solución sean independientes de la dimensión de los espacios en los cuales viven los datos \mathcal{L} (dimensionalmente pequeño) y \mathcal{H} (dimensionalmente grande).

Las máquinas de vectores soporte (SVM) se refieren a \mathcal{H} como un espacio de Hilbert.

Condición de Mercer

Se va a estudiar para qué kernels existe un par $\{\mathcal{H}, \phi\}$ con las propiedades descritas anteriormente y para cuales no. Las propiedades se cumplirán para los kernels que cumplan la condición de Mercer [31] [32]. Existe una aplicación ϕ y una expansión:

$$K(x, y) = \sum_{i=1} \phi(x)_i \phi(y)_i \quad (\text{B.2})$$

Si y solo si: Para cualquier $g(x)$, tal que

$$\int g(x)^2 dx$$

es finita, entonces

$$\int K(x, y) \cdot g(x) \cdot g(y) dx dy \geq 0$$

Se puede probar que cualquier kernel expresado de la forma: $K(x, y) = \sum_{p=0}^{\infty} c_p (x \cdot y)^p$, donde c_p son coeficientes positivos reales y la serie es uniformemente convergente, satisface la condición de Mercer.

Kernels de SVMs no lineales

Los primeros kernels investigados para el problema de reconocimiento de modelos fueron los siguientes:

$K(x, y) = (x \cdot y + 1)^p \rightarrow$ polinomio de grado p .

$K(x, y) = \exp -\frac{\|x-y\|^2}{2\sigma^2} \rightarrow$ clasificador de base radial gaussiana.

$K(x, y) = \tanh(kx \cdot y - \delta) \rightarrow$ kernel sigmoidal, da una red neuronal de dos capas.

Existen muchos otros tipos de kernels mucho más complejos que se pueden estudiar en la bibliografía ([33], [34], [35], [36], [37]).

Aunque los clasificadores SVM tratados hasta ahora han sido binarios, son fácilmente combinados para tratar múltiples clases.

Máquinas de vectores soporte no lineales para regresión

La solución para las máquinas de vectores soporte lineales para aplicarlas a regresión es:

$$\omega = \sum_{i=1}^N (\alpha_i - \alpha_{i'}) x_i \quad (\text{B.3})$$

Si se aplica al caso no lineal:

$$\omega = \sum_{i=1}^N (\alpha_i - \alpha_{i'}) \phi(x_i) \quad (\text{B.4})$$

Siguiendo el mismo procedimiento que el aplicado a las SVMs para problemas de clasificación, se puede encontrar la expresión para las SVRs no lineales:

$$y_j = \sum_{i=1}^N (\alpha_i - \alpha_{i'}) \phi(x_i)^T \phi(x_j) + b = \sum_{i=1}^N (\alpha_i - \alpha_{i'}) K(x_i, x_j) + b \quad (\text{B.5})$$

PROYECCIONES DE ENTRADAS EN EL ESPACIO DE CARACTERÍSTICAS

En la regresión lineal vimos que el modelo lineal Bayesiano sufre de expresividad limitada. Una idea muy sencilla para superar este problema es proyectar primero las entradas en algún espacio de alta dimensión usando un conjunto de funciones básicas y luego aplicar el modelo lineal en este espacio en lugar de directamente sobre las entradas.

Esta idea es usada en clasificación, donde un conjunto de datos que no son linealmente separables en el espacio de datos original pueden convertirse en linealmente separables en un espacio de características de alta dimensión.

Introducimos la función $\phi(x)$ que mapea un vector de entrada x de dimensión D en un espacio de características de dimensión N . Así definimos la matriz $\phi(X)$ como la agregación de columnas de $\phi(x)$ para todos los casos en el conjunto de entrenamiento. Ahora el modelo es:

$$f(x) = \phi(x)^T w, \quad (C.1)$$

donde ahora el vector de parámetros tiene longitud N . El análisis de este modelo es análogo al modelo lineal estándar, excepto que en todos los sitios X será sustituida por $\phi(X)$. De esta forma, la distribución predictiva es:

$$f_* | x_*, X, y \sim N\left(\frac{1}{\sigma_n^2} \phi(x_*)^T A^{-1} \phi y, \phi(x_*)^T A^{-1} \phi(x_*)\right). \quad (C.2)$$

donde $\phi = \phi(X)$ y $A = \sigma_n^{-2} \phi \phi^T + \Sigma_p^{-1}$. Sin embargo, podemos reescribir la ecuación de la siguiente manera:

$$f_* | x_*, X, y \sim N \left((\phi_*^T \Sigma_p \phi (K + \sigma_n^2 I)^{-1} y, \right. \\ \left. \phi_*^T \Sigma_p \phi_* - \phi_*^T \Sigma_p \phi (K + \sigma_n^2 I)^{-1} \phi^T \Sigma_p \phi_* \right), \quad (C.3)$$

donde hemos usado la taquigrafía $\phi(x_*) = \phi_*$, y hemos definido $K = \phi^T \Sigma_p \phi$.

En la ecuación C.4, el espacio de características siempre participa en la forma de $\phi^T \Sigma_p \phi^T$, $\phi_*^T \Sigma_p \phi$, $\phi_*^T \Sigma_p \phi_*$; así las entradas de estas matrices son invariables de la forma $\phi(x)^T \Sigma_p \phi(x')$, donde x y x' son los conjuntos de entrenamiento o de test.

Definimos $k(x, x') = \phi(x)^T \Sigma_p \phi(x')$. Nosotros llamaremos a esta función *función de covarianza o kernel*. $\phi(x)^T \Sigma_p \phi(x')$ es un producto interior (con respecto a Σ_p). Cuando Σ_p es definido positivo, podemos definir $\Sigma_p^{1/2}$ de tal forma que $(\Sigma_p^{1/2})^2 = \Sigma_p$.

Si un algoritmo se define únicamente en términos de productos interiores en el espacio de entrada, entonces puede ser suprimido en el espacio de características sustituyendo las apariciones de estos productos interiores por $k(x, x')$; esto se llama el *kernel trick*. Esta técnica es valuable en situaciones donde es más conveniente calcular el kernel que los propios vectores de características.

BIBLIOGRAFÍA

- [1] Roy CS, Sherrington CS. On the regulation of the blood-supply of the brain. *J Physiol.* 1890;11:85-108.
- [2] Clarke DD, Sokoloff L. Circulation and energy metabolism of the brain. En: Siegel GJ, Agranoff BW, *Basic Neurochemistry*. New York: Raven; 1994.p. 645-80.
- [3] Gross PM, Sposito NM, Pettersen SE, Panton DG, Fenstermacher JD. Topography of capillary density, glucose metabolism, and microvascular function within the rat inferior colliculus. *J Cereb Blood Flow Metab.* 1987;7:154-60.
- [4] Klein B, Kuschinsky W, Schrock H, Vetterlein F. Interdependency of local capillary density, blood flow, and metabolism in rat brains. *Am J Physiol.* 1986;251:H1333-H40.
- [5] Belliveau JW, Rosen BR, Kanton HL, Rzedzian RR, Kennedy DN, McKinstry RC, et al. Functional cerebral imaging by susceptibility-contrast NMR. *Magn Reson Med.* 1990;14:538-46.
- [6] Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, et al. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U.S.A.* 1992;89:5951-5.
- [7] Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, et al. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc Natl Acad Sci U.S.A.* 1992;89:5675-9.
- [8] Ríos M, Calvo Merino B, Álvarez Linera J. Tareas en Resonancia Magnética Funcional: Diseños de bloques, ligados a eventos y mixtos. En: VV.AA. *Avances en Neuropsicología clínica*. Madrid: Fundación MAPFRE; 2006.p. 163-96.
- [9] Donders FC. On the speed of mental processes. *Acta Psychologica.* 1969;30:412-31.

- [10] Posner MI, Petersen SE, Fox PT, Raichle ME. Localization of cognitive operations in the human brain. *Science*. 1988;240:1627-31.
- [11] Gusnard DA, Raichle ME. Searching for a baseline: Functional Imaging and the resting human brain. *Nat Rev Neurosci*. 2001;2:685-94.
- [12] Mazoyer B, Zago L, Mellet E, Bricogne S, Etard O, Houde O, et al. Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res Bull*. 2001;54:287-98.
- [13] Shulman GL, Corbetta M, Buckner RL, Raichle ME, Fiez JA, Miezin FM, et al. Top-down modulation of early sensory cortex. *Cereb Cortex*. 1997;7:193-206.
- [14] Grafton S, Mazziotta JC, Presty S, Friston K, Frackowiak R, Phelps M. Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *Journal of Neuroscience*. 1992;12:2542-8.
- [15] Álvarez-Linera J, Rios-Lago M, Martín H, Hernández-Tamames JA, Escribano J, Sánchez del Río M. Resonancia Magnética Funcional de la corteza visual: Estudio de las relaciones entre la intensidad del estímulo y la respuesta BOLD. *Rev Neurol*. En prensa.
- [16] Friston K, Price CJ, Fletcher PC, Moore C, Frackowiak RSJ, Dolan RJ. The trouble with cognitive subtraction. *Neuroimage*. 1996;4:97-104.
- [17] Raichle ME. A brief history of functional brain mapping. En: Toga AW, Mazziotta JC. *Brain Mapping: The systems*. San Diego, CA: Academic Press;2000.
- [18] V.Vapnik. *Statistical Learning Theory, Adaptative and Learning Systems for Signal Processing, Communications, and Control*. Simon Haykin, 1998.
- [19] F.Pérez-Cruz, P.Alarcón-Diana, A.Navia-Vázquez and A.Artés-Rodríguez. "Fast Training of Support Vector Classifiers. *Advances in Neural Information Processing Systems*". M.I.T.Press., November 2000.
- [20] Otten LJ, Henson RN, Rugg M. State-related and item-related neural correlates of successful memory encoding. *Nat Neurosci*. 2002;5:1339-44.
- [21] Chawla D, Rees G, Friston K. The physiological basis of attentional modulation in extrastriate visual areas. *Nat Neurosci*. 1999;2:671-6.
- [22] V.Vapnik. *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics*. New York, 1982.
- [23] V.Vapnik. *The nature of Statistical Learning Theory* Autor. New York, second edition, 2000.
- [24] V.Vapnik. Principles of risk minimization for learning theory. In M.Kaufmann, editor, *NIPS*, pages 831-838,1991.

- [25] B. Schölkopf. Support Vector Learning. R. Oldenbourg Verlag, Munich, 1997.
- [26] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):1-32, 1998.
- [27] M. Held, P. Wolfe, and H. P. Crowder. "Validation of subgradient optimization". *Mathematical Programming*, 6(1):62-88, 1974.
- [28] A. Smola and B. Schölkopf. A tutorial on support vector regression. 1988. *NeuroCOLT Technical Report NC-TR-98-030*, Royal Holloway College, University of London, UK.
- [29] R. Herbrich. *Learning Kernel Classifiers: Theory and Algorithms (Adaptive Computation and Machine Learning S.)*. The M.I.T Press, Cambridge, Massachusetts, 2002.
- [30] S. Mika, G. Ratsch, J. Weston, B. Schölkopf, A. J. Smola, and K. R. Müller. Invariant feature extraction and classification in kernel spaces. In *Advances in Neural Information Processing Systems*. The M.I.T Press, 12:526-532, 1999. Cambridge, MA.
- [31] R. Courant and D. Hilbert. "Methods of Mathematical Physics". John Wiley and Sons, 1953.
- [32] M. A. Aizerman, E. M. Braverman, and L. I. Rozonoer. "Theoretical foundations of the potential function method in pattern recognition learning". *Automation and Remote Control*, 25:821 to 837, 1964.
- [33] A. Zien, G. Ratsch, S. Mika, B. Schölkopf, T. Lengauer, and K. R. Müller. Engineering support vector machine kernels that recognize translation initiation sites. *Bioinformatics*, 16(9):799-807, 2000.
- [34] D. Haussler. Convolution kernels on discrete structure. University of California at Santa Cruz, July 1999. Santa Cruz, CA, USA, Tech. Rep. UCSC-CRL-99-10.
- [35] T. Jaakkola, M. Diekhans, and D. Haussler. A discriminative framework for detecting remote protein homologies. *Journal of Computational Biology*, 7(1,2):95-114, 2000.
- [36] A. J. Smola, B. Schölkopf, and K. R. Müller. "The connection between regularization operators and support vector kernels". *Neural Networks*, 11(4):637-649, 1998.
- [37] K. Tsuda, M. Kawanabe, G. Rätsch, S. Sonnenburg, and K. R. Müller. A new discriminative kernel from probabilistic models. *Neural Computation*, 14(10):2397-2414, 2002.
- [38] M. Brett. The MNI brain and the talairach atlas, February 2002.

- [39] W. Chau and A. R. McIntosh. “The Talairach coordinate of a point the MNI space: how to interpret it”. *Neuroimage*, 25(2):408-416, April 2005.